

Discrete Gauge Symmetries in Discrete MSSM-like Orientifolds

L.E. Ibáñez^{1,3} A.N. Schellekens² and A. M. Uranga³

¹ Departamento de Física Teórica, Módulo 8, Facultad de Ciencias,
Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

² Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands;
IMAPP, Radboud Universiteit, Nijmegen, The Netherlands;
Instituto de Física Fundamental (IFF-CSIC),
Serrano 123, Madrid 28006, Spain.

³ Instituto de Física Teórica UAM/CSIC,
c/ Nicolás Cabrera 13-15, 28049 Madrid, Spain

Abstract

Motivated by the necessity of discrete \mathbb{Z}_N symmetries in the MSSM to insure baryon stability, we study the origin of discrete gauge symmetries from open string sector $U(1)$'s in orientifolds based on rational conformal field theory. By means of an explicit construction, we find an integral basis for the couplings of axions and $U(1)$ factors for all simple current MIPFs and orientifolds of all 168 Gepner models, a total of 32990 distinct cases. We discuss how the presence of discrete symmetries surviving as a subgroup of broken $U(1)$'s can be derived using this basis. We apply this procedure to models with MSSM chiral spectrum, concretely to all known $U(3) \times U(2) \times U(1) \times U(1)$ and $U(3) \times Sp(2) \times U(1) \times U(1)$ configurations with chiral bi-fundamentals, but no chiral tensors, as well as some $SU(5)$ GUT models. We find examples of models with \mathbb{Z}_2 (R-parity) and \mathbb{Z}_3 symmetries that forbid certain B and/or L violating MSSM couplings. Their presence is however relatively rare, at the level of a few percent of all cases.

1 Introduction

Discrete symmetries are often used in model building in order to argue that certain otherwise allowed terms in the effective Lagrangian are in fact absent. So far in nature no such symmetry has been observed, apart from CPT. Furthermore, all allowed standard model interactions are indeed observed, with the exception of the QCD strong CP violating term, whose apparent absence can however not be explained in terms of an exact discrete symmetry.

On the other hand, it has been quite common to invoke discrete symmetries in model building beyond the standard model, and especially supersymmetric model building. Indeed, a generic point in the full parameter space of the supersymmetrized standard model with soft supersymmetry breaking terms would be in disagreement with observation, most notably because of the allowed baryon and/or lepton number violating UDD , QLD , LLE and LH_u terms¹ leading to dimension four operators. These give rise to catastrophic proton decay rates if all of them are present with coefficients of order one, and serious constraints even if some of them are absent. The most common solution to this problem is to postulate a discrete symmetry that forbids them, such as R-parity [1] or Baryon-triality [2].

Even if the dimension four operators are absent, one has to worry² about dimension five operators originating from the superpotential terms $QQQL$ and $UUDE$, which preserve $B-L$ and are therefore not forbidden by R-parity. This is because R-parity may be considered as a \mathbb{Z}_2 subgroup of $B-L$.

In string theory one cannot simply postulate a discrete symmetry. It must be a verifiable property of a given string realization. There are strong arguments supporting that global symmetries (either continuous or discrete) cannot exist in theories containing quantum gravity, rather they must be gauge symmetries (see e.g. [5] for a recent discussion, and [6, 7, 8, 9, 10, 11, 12] for earlier references). Hence, any exact discrete symmetry in string theory must be gauge. In chiral models, discrete symmetries are strongly constrained by anomaly cancellation conditions [13] (see also [14, 15]). For MSSM spectra there is a short list of allowed possibilities [2] (see also [16] and references therein; see also [17, 18, 19, 20, 21]).

¹Here Q, U, D, L and E denote the usual MSSM superfields, with U, D and E referring to anti-particles. The two Higgs superfields are denoted H_u and H_d , and couple to up and down quarks respectively.

²These operators would be less problematic if proton decay rates are suppressed non-perturbatively, as suggested recently in [3]. However, lattice computations do not show such a suppression, see e.g. [4] and references therein.

Abelian \mathbb{Z}_n discrete gauge symmetries can be usefully realized as discrete remnants of $U(1)$ gauge symmetries with BF couplings. This structure occurs frequently for open string sector $U(1)$'s in orientifold models (see [22] for a study in geometric orientifold compactifications). These models are obtained by starting from a closed type-II superstring theory, modding out a world-sheet orientation reversing symmetry, and adding an open string sector, with the open strings ending on a brane (see [23] for review and references). The Chan-Paton group of each brane is either $O(N)$, $Sp(N)$ or $U(N)$. In the latter case there is always a $U(1)$ gauge symmetry, which is often anomalous. If so, the anomaly is cancelled by a Green-Schwarz mechanism involving one of the RR axions, of which there are usually many. In this process the $U(1)$ gauge boson acquires a mass through a BF couplings, equivalently, by absorbing a scalar field (henceforth dubbed ‘axion’) as its longitudinal component. Furthermore, even non-anomalous $U(1)$'s can acquire a mass through this kind of BF interaction [24].

The perturbative couplings always respect charge conservation with respect to any of these $U(1)$'s, anomalous or not, massive or massless. For example, in brane models where baryon number and lepton number are embedded in these brane $U(1)$'s, no B or L violating couplings are generated perturbatively. However, massive $U(1)$'s are in general violated by non-perturbative effects, mediated by brane instantons coupling to the relevant axion field. The branes supporting the instanton may be part of the brane configuration of the model, in which case they correspond to gauge instantons, and have strength controlled by gauge couplings, hence negligible in SM-like models. More interestingly, the instantons can originate from different branes, and have strength unrelated to gauge couplings, and hence can potentially induce sizable values for the coefficients for the $U(1)$ violating operators. Examples of such “stringy instantons” have been discussed in many papers [25, 26, 27, 28, 29, 30, 31, 32, 33, 34]. Such instantons may in general have a surplus of zero-modes, so that they do not contribute to the superpotential, but the violation of the symmetry arises in higher derivative terms [35, 36, 37]. It is thus important to characterize the general set of $U(1)$ violating instantons in any given string model.

This is in general difficult to achieve by explicit construction. If a brane instanton is found with suitable zero-mode structure to induce a coupling, this proves that the latter is generated; but if no such explicit instanton can be found, it does not follow that the coupling is *not* generated. This is especially true in RCFT constructions, where *a priori* only a limited set of branes (boundary states) is available. This is where discrete symmetries can play a useful rôle. It may happen that a $U(1)$ is not broken

completely, but that a discrete \mathbb{Z}_n symmetry remains. In that case, no instanton, and indeed no no-perturbative effect can break it, since it is an exact symmetry of the theory.

A second motivation to study the existence of discrete gauge symmetries is that, apart from forbidding undesirable terms, discrete symmetries may also forbid desirable ones, such as certain Yukawa couplings, or Majorana mass terms for right-handed neutrinos. Hence, the study of discrete symmetries can help one to avoid the pointless exercise of looking for instantons that cannot exist anyway, or to focus on those models which potentially allow for them.

Finally, one may also ask a different kind of question. Rather than determining if a given model does or does not have a certain discrete symmetry, one may ask if the occurrence of such symmetries is a generic phenomenon in string theory, or at least in certain subclasses. If the answer is positive, an appeal to discrete symmetries to avoid catastrophic couplings becomes *more* credible, but then it becomes *less* convincing that e.g. Majorana mass terms or perturbatively forbidden Yukawas will generically be generated by instantons.

In this respect, we note that our results on the frequency of occurrence of discrete symmetries are not directly related to other work on the occurrence of discrete R-symmetries in the string landscape, such as [38, 39, 40]. The latter deal with closed type-II strings (including some type-II Gepner models), without open string sectors, whereas we study symmetries originating from Chan-Paton groups of open strings. Also, the discrete symmetries discussed in the present paper act on all members of supermultiplets in the same way, and are therefore *not* discrete R-symmetries. Nevertheless, our discussion *does* include R-parity, since despite its name it can be obtained as the discrete \mathbb{Z}_2 subgroup of $B-L$ generated by $(-1)^{3(B-L)}$ (physically indistinguishable from the \mathbb{Z}_2 R-symmetry $(-1)^{3(B-L)+2S}$, since in any scattering amplitude S is conserved modulo integers).

For all the reasons explained above it is important to be able to compute in an efficient way and in a large classes of models if these orientifold $U(1)$'s are completely or only partly broken. This is what we wish to do here for the class of Gepner orientifolds (orientifolds of the closed type-II string theories first constructed in [41]).

Gepner orientifolds models [42, 43, 44, 45, 46, 47, 48, 49, 50, 51] provide access to an interesting region of the landscape where open string configurations can be found that realize the standard model. The methods used are quite different than the more familiar ones used for orientifolds of orbifolds or Calabi-Yau models. The basic ingredients are

not branes wrapping cycles on a manifold, but boundary states in rational conformal field theory. However most of the techniques available in brane descriptions can be translated rather easily to rational CFT language. For example, it is known how to compute massless (and even massive) spectra corresponding to brane intersections, how to cancel disk tadpoles against crosscap tadpoles, and how to check if a $U(1)$ is broken by couplings to axions. Examples exist [47] where the resulting unbroken gauge group is *exactly* $SU(3) \times SU(2) \times U(1)$. In this example all tadpoles cancel within the standard model sector, hence there are no “hidden sector” gauge groups, and all superfluous continuous $U(1)$ ’s, especially the usually problematic $B-L$ symmetry, are broken.

In [22] it was explained how to find $U(1)$ ’s broken to discrete subgroups in the geometric approach. We would like to translate this result to RCFT, or at least the special case of Gepner models. Here we immediately run into a problem. In the geometric setting, discrete symmetries can be read off from the axion-gauge boson couplings, which in a suitable geometric basis are integers. The signature of a discrete symmetry is a common factor of these integers. However, in a RCFT setting, the boundary coefficients are complex numbers, and a canonical basis in which they are integers is not readily available.

In this paper we solve this problem in an empirical way, by developing an algorithm that does allow us to write all coefficients in terms of integers. This algorithm yields a preferred set of boundary states that plays the same rôle as the aforementioned geometric basis. This involves extensive numerical computations, which give the desired result in all cases. The fact that this is possible calls for a deeper understanding, a principle that determines the basis without extensive computations. This in its turn may provide new insights in the geometric interpretation of all these orientifold models. However, for our present purpose the empirically determined basis does the job. We hope to return to the underlying structure in the future.

This paper is organized as follows. In the next section we review some basic details of Gepner Models. In section 3 we define the problem and explain how it is solved. In section 4 we explain the algorithm that yields the integral basis. In section 5 we give some examples. Some conclusions are offered in section 6.

2 Gepner orientifolds

A given Gepner model is specified by a tensor product of $N=2$ minimal models and a modular invariant partition function (MIPF for short). The tensor product has total central charge 9. There are 168 ways of tensoring minimal $N=2$ models so that a the required central charge 9 is obtained. They can be denoted as (k_1, \dots, k_N) , where $k_i \geq 1$ is the level of factor i . Usually we drop the commas and the brackets, and denote a tensor product by a numerical string built out of the k_i in ascending order. This notation turns out to be unambiguous. Each of the 168 Gepner models has a chiral algebra \mathcal{A} , containing the $N=2$ super-Virasoro algebra. This algebra has a finite number (typically of order $10^3 \dots 10^5$) of representations, labelled by a set of integers. To get these representations, one starts with all combinations of representations of the N tensor factors, and subjects them to a projection to impose world-sheet supersymmetry. In addition to the N tensor factors, the four-dimensional NSR fermions are also participating in a non-trivial way. In addition to this, one extends the chiral algebra with a spin-1 operator which is a space-time spinor and imposes a GSO projection. The result of these two extensions is the algebra \mathcal{A} . Here and in the following we use the simple current [52, 53] description of MIPFs of Gepner models presented in [54]. It has the advantage that all world-sheet and space-time supersymmetry projections can be treated on equal footing with the construction of the MIPFs, and that explicit details about the $N=2$ primaries can be omitted. All we need to know is how the simple currents act on them.

The most general simple current MIPFs for each of the 168 Gepner models can be built using the formalism developed in [55]. This gives rise to a total of 5392 distinct MIPFs. They are characterized by a non-negative integer multiplicity matrix M_{ij} indicating how often certain left-right character combinations occur in the closed string. The labels i refer to the aforementioned set of chiral algebra representations. Permutations of identical $N=2$ factors³ generate isomorphisms between MIPFs. These have been removed, so that with a few exceptions all MIPFs are really distinct. There are just a few cases where two or more supposedly distinct MIPFs yield apparently

³In the context of RCFT orientifolds, a naively possible origin of discrete symmetries is the permutations of identical factors. However, many of these permutations are broken by the MIPF, orientifold and boundary state choice. Although it is possible that cases exist where permutation symmetries survive in the spectrum as exact discrete symmetries, we do not know any examples. In any case this is not the subject of the present paper, which as already mentioned focuses on symmetries arising from open string sector $U(1)$'s.

identical data.

On top of this there are choices to define the orientifold quotient [58]. Taking all of these into account brings the total number of possibilities to about 49000. However, some of them have zero tension, which implies that no branes can be added without violating the dilaton tadpole condition. After eliminating these cases we end up with 32990 in principle distinct orientifolds (as with the MIPFs, in practice there are always a few “accidental” degeneracies that are apparent in the spectrum, but do not have a very obvious fundamental origin. This is irrelevant in practice).

A simple current MIPF is characterized by a discrete group \mathcal{H} of simple currents, and a matrix of rational numbers $X(M, J)$ defined on \mathcal{H} . On Riemann surfaces with boundaries each MIPF has a definite set of Ishibashi states and a corresponding set of boundary states. The former are simply in one-to-one correspondence with the elements M_{ii^c} of the multiplicity matrix, where i^c denotes the two-dimensional charge conjugate of i . In a simple current MIPF these states are labelled by a label m referring to a representation of the chiral algebra of the tensor product, and a degeneracy label J . For each m , this degeneracy label is the simple current in \mathcal{H} that fixes it, *i.e.* $Jm = m$, with $M_{mm^c} \neq 0$. So Ishibashi states will be denoted as (m, J) .

The set of boundary states that respects all the symmetries of the original chiral algebra is known to be equal to the number of Ishibashi states [56]. They are characterized by the orbits of \mathcal{H} on the chiral algebra representations. These orbits can be labelled by an integer a that belongs to the set of representation labels of the full chiral algebra. An orbit is a set of representation labels related by the action of \mathcal{H} . For the boundary label we choose one representative from this set. Also in this case there may be degeneracies, which occur if the \mathcal{H} -action has fixed points. The degeneracy labels can be conveniently chosen as the discrete group character ψ of certain subgroup (called the “central stabilizer”) \mathcal{C}_a of the stabilizer \mathcal{S}_a of a (the stabilizer is the subgroup of \mathcal{H} of that fixes a representation a). The boundary labels are then $[a, \psi_a]$. Note that the set of characters depends on the boundary label. If the central stabilizer is a discrete group with $|\mathcal{C}_a|$ elements, than there exists exactly $|\mathcal{C}_a|$ distinct characters (complex functions on \mathcal{C}_a that respect the group property).

Now we have two sets (m, J) and $[a, \psi_a]$ of Ishibashi and boundary labels. These can be shown to be of equal size, although this is not manifest. On this basis we now define boundary reflection coefficients [58]

$$R_{[a, \psi_a](m, J)} = \sqrt{\frac{|\mathcal{H}|}{|\mathcal{C}_a||\mathcal{S}_a|}} \psi_a^*(J) S_{am}^J \quad (2.1)$$

Here S_{am}^J is a matrix element of the modular transformation matrix of a certain algebra associated with the original chiral algebra and the current J [59, 60]. If J is the identity, S^J is equal to the modular transformation matrix of the chiral algebra \mathcal{X} . If J is not the identity, S^J is the modular transformation matrix of another algebra, usually, but not always, related to some other conformal field theory. All these matrices are explicitly known and are, in general, complex numbers. In the prefactor $|\mathcal{H}|$, etc, denotes the number of elements of the corresponding discrete group.

The boundary coefficients are independent of the orientifold choice. The latter enters the discussion in two ways. First of all the unoriented annulus coefficients have the form [58]

$$A_{[a,\psi_a][b,\psi_b]}^i = \sum_{m,J,J'} \frac{S_m^i R_{[a,\psi_a](m,J)} g_{J,J'}^{\Omega,m} R_{[b,\psi_b](m,J')}}{S_{0m}} \quad (2.2)$$

Here $g_{J,J'}^{\Omega,m}$ is an orientation-dependent metric on the space of Ishibashi states; Ω denotes the orientifold choice. In general, $g_{J,J'}^{\Omega,m}$ is a block-diagonal matrix in the label m , which can act non-trivially in the degeneracy spaces for each m . One could in principle take the square root of this metric and absorb it into the boundary coefficients, which then become orientation dependent. However, it is both physically more appealing and also more convenient to have orientifold-independent boundary coefficients. The final results will not be affected by this convention.

To make the notation a bit less cumbersome we will use in the following a single letter “ a ” instead of the combination $[a, \psi_a]$ to denote boundaries. The fixed point splitting of the boundary labels does not really play a rôle in what follows.

The annulus coefficients appear in the expression for the oriented annulus as

$$A = \sum_{a,b} N_a N_b \sum_i A_{ab}^i \chi^i(\tau/2) , \quad (2.3)$$

where N_a, N_b are the Chan-Paton multiplicities, and χ^i are the Virasoro characters of the chiral algebra \mathcal{X} .

The second way the orientifold choice matters is in boundary conjugation. This is defined as follows

$$A_{ab}^0 = \begin{cases} 1 & \text{if } b = a^c \\ 0 & \text{otherwise} \end{cases} \quad (2.4)$$

Clearly the dependence of boundary conjugation on orientation can be traced back to the Ishibashi metric $g_{J,J'}^{\Omega,m}$, so that in the end all dependence on orientation can be traced back to this quantity.

3 Axion couplings

3.1 Discrete \mathbb{Z}_N symmetries from open string $U(1)$'s

The key to understand the appearance of discrete \mathbb{Z}_N gauge symmetries from open string $U(1)$'s are the coefficients R_{am} , which determine the BF couplings.

Consider a 4d string model, with a set of branes labelled with a and their orientifold images a^c , with BF couplings to a set of RR 2-forms B_m

$$\sum_{a,m} N_a V_{am} B_m \wedge F_a \quad (3.1)$$

Here $V_{am} = R_{am} - R_{a^c m}$, with the relative minus sign arising because the physical $U(1)$ gauge boson is the difference of those supported on the brane and its orientifold image.

Consider now a linear combination⁴ $\sum_a x_a Y_a$ of the $U(1)$ generators Y_a of brane a . Its BF couplings are

$$\sum_m \left(\sum_a x_a N_a V_{am} \right) B_m \wedge F \quad (3.2)$$

It thus remains massless if and only if

$$\sum_a x_a N_a (R_{am} - R_{a^c m}) = 0 \text{ for all } m. \quad (3.3)$$

In general, the set of massless $U(1)$'s correspond to the space of zero eigenvectors x_a of the non-symmetric matrix $M_{am} = N_a (R_{am} - R_{a^c m})$.

Massive $U(1)$'s are broken by brane instantons coupling to the axion RR scalars ϕ_m dual to the 2-forms. With a suitable normalization, the amplitudes go like $e^{-2\pi i \phi_m}$, and the axions have an identification $\phi_m \simeq \phi_m + 1$. It is useful to introduce the dual description of (3.2) in terms of ϕ_m . The relevant lagrangian is

$$\sum_m [\partial_\mu \phi_m - (\sum_a x_a N_a V_{am}) A_\mu]^2 \quad (3.4)$$

where the $U(1)$ is normalized such that the minimal charge is 1. Under $U(1)$ transformations,

$$A_\mu \rightarrow A_\mu + \partial_\mu \lambda \quad ; \quad \phi_m \rightarrow \phi_m + (\sum_a x_a N_a V_{am}) \lambda \quad (3.5)$$

Instanton amplitudes transform as

$$e^{-2\pi i \phi_m} \rightarrow e^{-2\pi i \phi_m} \exp[-2\pi i (\sum_a x_a N_a V_{am}) \lambda] \quad (3.6)$$

⁴It is useful to maintain the convenient normalization that $U(1)$'s have minimal charge 1; this requires the x_a to be integer, with $\gcd(x_a) = 1$.

and this transformation is cancelled by the insertion of an operator in the charged matter, with total charge $(\sum_a x_a N_a V_{am})$. This quantity therefore measures the amount of $U(1)$ violation.

Thus the condition for a discrete \mathbb{Z}_N remnant of the $U(1)$ is therefore

$$\sum_a x_a N_a (R_{am} - R_{a^c m}) = 0 \bmod N \quad \text{for all } m. \quad (3.7)$$

This \mathbb{Z}_N is an exact gauge symmetry. The result also follows from the analogy of (3.4) with the Higgsing of the $U(1)$ by a charge N scalar (whose phase is played by a suitable linear combination of the RR scalars).

In other words, to find \mathbb{Z}_N discrete symmetries, we should look for zero eigenvectors modulo N of the matrix M_{am} , in the convenient normalization used above. In a geometric setting, they can be made integer by choosing a suitable basis for the axions, in terms of basic 3-cycles on the compactification manifold (in type-IIA language), see [22]. But this notion (and so the automatic appearance of the convenient normalization) is not readily available for Gepner models, although in some very special cases (the quintic Calabi-Yau) similar bases have been discussed [57]. A specialized geometric discussion for each separate case is not likely to get us to the desired result, since we will have to deal with all possible simple current MIPFs of the 168 Gepner models, and all their orientifolds, a total of 32990 distinct possibilities with non-vanishing orientifold tension.

The general formula for the boundary coefficients in such a CFT takes the form (2.1). Here S_{am}^J is a modular transformation matrix of a conformal field theory, and ψ is a phase, and the pre-factor is a square root of a rational number. Neither of these factors are integers. Indeed, in general these boundary coefficients are complex. It is not clear how to even define condition (3.7). The key towards the resolution, is to search for a basic set of instanton branes, whose boundary coefficients thus define the axion periodicities; this then allows to effectively move onto the convenient normalization in which coefficients are integers, whose gcd then gives the order of the discrete symmetry.

3.2 Axions in RCFT and basis of boundary states

In this section we describe the structure of axions in RCFT orientifolds, and explain the relevance of the above mentioned basis of boundary states in RCFT terms.

An Ishibashi state (m, J) contains an axion if the representation m contains a massless space-time spinor. The ground state may contain N_L left-handed and N_R right-handed massless spinors. In the closed string one gets the square of the character

multiplied by the multiplicity matrix M_{ij} . This is then subject to the Klein bottle projection if $i = j$ (since the Ishibashi states correspond to $i = j^c$, the Klein bottle projection can only affect self-conjugate Ishibashi states with $N_L = N_R$; for a more detailed discussion see the appendix). However all these multiplicities are ignored in the following, because they all have the same boundary coefficients. Each representation with $N_L + N_R > 0$ is counted as one axion, regardless of the values of these numbers and Klein bottle projected closed string multiplicity and even if the Witten index $N_L - N_R$ vanishes. Note however that each degeneracy label J is counted once as a different axion, since the axion coefficients depend on J .

We introduce the following notation for the relevant coefficients.

$$V_{a\nu} \equiv R_{a(m,J)} - R_{a^c(m,J)} , \quad (3.8)$$

where $a = 1, \dots, N_B$, the number of complex boundary pairs. The second label ν identifies contributing axions according to the rules stated above. We will discuss the precise range of the label below. Charge conjugation is defined by means of the orientifold choice. We regard these objects as a complex matrix with columns labelled by ν and rows by a . Note that self-conjugate boundaries do not contribute, and that each complex pair contribute one row, so that there is a one-to-one correspondence between the rows and all possible $U(1)$ factors in the open string spectrum.

As stated above, these coefficients $V_{a\nu}$ are in general complex. However, since they are coefficients of BF couplings, they should be real (morally, up to phase redefinitions of the RR fields); in other words, they can be made real by an a -independent but ν -dependent phase rotation (as can be shown directly in explicit models, see next section), as we assume in the following.

Note however that, since the axion periodicities had not been fixed to unity, they are not integers or rational numbers. However we will demonstrate in the next section that there exists at least one choice of boundaries $c(\mu)$ such that the following relation holds

$$V_{a\nu} = \sum_{\mu=1}^{N_A} Q_{a\mu} V_{c(\mu)\nu} , \quad Q_{a\mu} \in \mathbb{Z} \quad (3.9)$$

Here c is a map from the set of axion labels into the set of boundary labels which assigns a different boundary label to each axion label: $c(\mu) \neq c(\nu)$ if $\mu \neq \nu$. The number of basis vectors N_A is equal to the number of independent columns of $V_{a\nu}$ (it turns out to be sufficient to remove vanishing and identical columns, more complicated dependencies do not occur). Consequently, the labels ν cannot be expressed unambiguously in terms

of the original transverse channel labels (m, J) : there may be more than one (m, J) corresponding to any given ν . More details will be given in the next section.

If (3.9) can indeed be realized, it defines a basis in the space of all complex boundaries such that all other boundaries can be expanded in that basis with integer coefficients. In this way we obtain a lattice of charges, so that each boundary state corresponds to a point on that lattice. In general we expect that this basis is not unique, just as the basis of a lattice is not unique. Note however that not every lattice point is occupied. This is obvious because there is only a finite number of boundary states and an infinite number of lattice points, but also near the origin there are in general unoccupied sites. This implies that not every lattice basis can be realized in terms of boundary states.

The basic boundary states defines a set of ‘smallest instantons’ (at least in the RCFT realm), whose couplings to the axions *define* the axion periodicities. The quantities $Q_{a\mu}$ thus correspond to the coefficients of the BF couplings in the desired normalization in which the axions have unit periodicity, and can therefore be used to look for the discrete \mathbb{Z}_N symmetry. Namely a $U(1)$ integer linear combination $Y = \sum_a x_a Y_a$ (with the conventions in footnote 4) has an unbroken \mathbb{Z}_N subgroup if it satisfies the condition

$$\sum_a x_a N_a Q_{a\mu} = 0 \pmod{N} \quad (3.10)$$

There is an alternative description of the physical relevance of the basis, which instead of leaning on the axion periodicities, is based on expressing the amount of instanton $U(1)$ violation in terms of the basic instantons, as follows (both viewpoints are clearly related since (3.5) links $U(1)$ gauge transformations and axion shifts). As described in [25, 26], the amount $I_b(a)$ of $U(1)_a$ violation by an instanton supported on a brane b is given by the net number of charged fermion zero modes arising from massless open strings stretching between both boundaries. In the RCFT setup, and accounting for orientifold images, we have a combination of the annulus coefficients (2.3)

$$I_b(a) = N_a \sum_i w_i (A_{ba}^i - A_{ba^c}^i) \quad (3.11)$$

where w_i is the Witten index in the open string sector, which effectively extracts the net chiral contribution. Using (2.2) we have

$$I_b(a) = N_a \sum_i w_i \sum_{m, J', J} \left[\frac{S_{im} R_{b(m, J')} g_{J'J}^{\Omega, m}}{S_{0m}} \right] (R_{a(m, J)} - R_{a^c(m, J)}) \quad (3.12)$$

Note that these quantities are integer, and moreover can be defined for any boundary states a, b , regardless of whether a actually realizes a $U(1)$ symmetry in the model or not. Decomposing the boundary coefficients using (3.9), and reconstructing back to annulus amplitudes, we obtain

$$I_b(a) = \sum_{\mu} N_a Q_{a\mu} I_b(c(\mu)) \quad (3.13)$$

Here $I_b(c(\mu))$ are formally defined as in (3.12); in physical terms, they are integers measuring the violation by the instanton brane b of a *putative* $U(1)$ carried by brane $c(\mu)$ (which need not support an actual $U(1)$ of the model). For a $U(1)$ linear combination $Y = \sum_a x_a Y_a$ (with the conventions in footnote 4), the charge violation by an instanton brane b is

$$I_b(x) = \sum_a x_a I_b(a) = \sum_{\mu} \left(\sum_a x_a N_a Q_{a\mu} \right) I_b(c(\mu)) \quad (3.14)$$

Since $I_b(c(\mu))$ are integer, if the coefficients $\sum_a x_a N_a Q_{a\mu}$ have a common factor N , all instantons violate $U(1)$ charge in multiples of N , so that a discrete \mathbb{Z}_N subgroup remains unbroken. Hence we recover condition (3.10) for the existence of a discrete \mathbb{Z}_N symmetry.

Although this derivation exploited the RCFT formulas, eq. (3.13) makes full physical sense even for non-RCFT instantons. This strongly supports that the result holds for any instanton b , and therefore that the proposed condition (3.10) is correct in general. Still, it is possible that the basic quantities $I_b(c(\mu))$ already have a common factor. If they do not, we will get a \mathbb{Z}_N discrete symmetry, as read off from the coefficients $Q_{a\mu}$; otherwise, we can only get more discrete symmetries than naively expected. We believe this possibility to be fairly unlikely. The fact that we were able to find an integral lattice of charges in all cases strongly suggests that (3.10) identifies the discrete symmetries correctly.

4 Finding an integral basis

We will now explain a method that turns out to be very effective to find the integral basis described above.

Our starting point is the matrix $V_{a\nu}$, where rows a label boundary states and columns ν label axion fields. First we will normalize the coefficients $V_{a\nu}$ in a convenient way. In their raw form, these coefficients are not even relatively real. However, on

already explained physical grounds, they can be made real with a independent phase redefinitions, which are duly accounted in the following normalization. Consider in each column the first non-vanishing entry, starting at the top. If there is such an entry, divide all entries in the column by it, so that the top entry is equal to 1. If there are any columns that are completely zero, we discard them, since they describe decoupled axions; also, if two columns are identical, we keep only one of the two, since there is a decoupled linear combination of axions. This procedure only eliminates vanishing or identical columns. This is in general not sufficient to ensure that the columns are linearly independent, although this turns out to be the case in practice in the whole class of models. We call the dimension of this axion space N_A . As expected, it turns out that after normalizing the top entry of each column to 1, all entries in the matrix become real numbers.

This normalization removes some convention-dependent factors in the boundary coefficients. For instance, as mentioned below (2.2), we could have defined the boundary coefficients differently by absorbing the square root of $g_{J,J'}^{\Omega,m}$ in them; this is conveniently done by choosing a basis in degeneracy space so that $g_{J,J'}^{\Omega,m}$ is diagonal, so it can be absorbed into the boundary coefficients by multiplying each column by a certain complex factor. The normalization procedure discussed in the foregoing paragraph removes any possible dependence on such conventions.

Note that this normalization procedure depends on the way the boundaries are ordered. This ordering is not just arbitrary, because it descends from the ordering of the representations of the chiral \mathcal{X} , but the ordering is not in any way canonical either. Roughly speaking, it has the property that if $i > j$, then $S_{0i} > S_{0j}$, but even that ordering is not strict. However, the final result will not depend on this normalization procedure.

In this way we now obtain a real matrix $V_{a\mu}$, where a labels boundaries and μ the reduced set of axionic Ishibashi states. Now we consider the inner product matrix

$$N_{ab} = \sum_{\mu} V_{a\mu} V_{b\mu} \equiv V_a \cdot V_b \quad (4.1)$$

Upon explicit computation, it turns out that this is a rational matrix in all models, even though the coefficients V are real, and in general not rational. Note that if we renormalize an entire column by \sqrt{q} , $q \in \mathbb{Z}$ this does not affect the rationality. However, it is not uncommon to encounter other irrational numbers such as $p + \sqrt{q}$ and sine and cosines of rational multiples of π . It is therefore far from obvious that the rationality of V will persist if we order the boundaries differently, thus obtaining a

different normalization prescription. However, it is an empirical fact that in all 32990 cases of different MIPFs and orientifolds all these numbers N_{ab} come out rational, and this will turn out to be a very fortunate outcome.

Based on the intuitions in earlier sections, the hope is to find a basis in the space of Ishibashi states such that all coefficients $V_{a\mu}$ are transformed into integers, *i.e.* find a real and invertible matrix R such that

$$Q_{a\nu} = \sum_{\mu} V_{a\mu} R_{\mu\nu} \in \mathbb{Z} \quad (4.2)$$

If such a basis exists, the coefficients $V_{a\mu}$ can be written as

$$V_{a\nu} = \sum_{\mu} Q_{a\mu} R_{\mu\nu}^{-1} \quad (4.3)$$

We may think of the matrix $R_{\mu\nu}^{-1}$ as a set of basis vectors $B_{\nu}^{(\mu)}$ labelled by μ , and then what we are looking for is a set of basis vectors in terms of which all vectors $V_{a\nu}$ have integer expansions. In other words, all vectors $V_{a\nu}$ lie on the lattice spanned by the basis vectors. If we express the inner products N_{ab} in (4.1) in terms of the basis vectors we get

$$N_{ab} = \sum_{\mu} \sum_{\nu} Q_{a\mu} Q_{b\nu} \sum_{\rho} R_{\mu\rho}^{-1} R_{\nu\rho}^{-1} = \sum_{\mu} \sum_{\nu} Q_{a\mu} Q_{b\nu} B^{\mu} \cdot B^{\nu} \quad (4.4)$$

This tells us that if the basis vectors have integer (or rational) inner products, then integrality (rationality) of *all* N_{ab} follows automatically.

It is then natural to conjecture that the basis vectors might themselves be chosen as a subset of the boundary vectors $V_{a\mu}$. A necessary condition is that we should be able to find N_A independent vectors $V_{a\mu}$. Here it is important that the N_A columns are linearly independent, as explained above. A basis of this kind is defined by a map $c(\mu)$ from the set of axion labels to the set of boundaries, and we write

$$R_{\mu\nu}^{-1} = B_{\nu}^{(\mu)} = V_{c(\mu)\nu} \quad (4.5)$$

After inverting this matrix we can compute the charges using (4.2). The fact that all N_{ab} are rational guarantees that the charges are rational. But we can do better than that. Suppose some boundary vector W has the following expansion in terms of the basis

$$W_{\nu} = \sum_{\mu} Q_{\mu} V_{c(\mu)\nu} = \sum_{\mu} \frac{p_{\mu}}{q_{\mu}} V_{c(\mu)\nu} , \quad (4.6)$$

where p_{μ} and q_{μ} are relative prime. Now suppose there is one μ , denoted $\hat{\mu}$, so that $p_{\hat{\mu}} = 1$. We may then bring the corresponding term to the left, W_{ν} to the right and

multiply by $-q_{\hat{\mu}}$. Then we get

$$V_{c(\hat{\mu})\nu} = \sum_{\mu, \mu \neq \hat{\mu}} -\frac{p_{\mu} q_{\hat{\mu}}}{q_{\mu}} V_{c(\mu)\nu} + q_{\hat{\mu}} W_{\nu} \quad (4.7)$$

Now we may remove $V_{c(\hat{\mu})}$ from the basis and replacing it by W , thus defining a new map, $\hat{c}(\nu)$. The advantage is that now one of the charges has changed from $1/q_{\hat{\mu}}$ to $q_{\hat{\mu}}$. Furthermore, if q_{μ} and $q_{\hat{\mu}}$ have common factors, the remaining denominators are reduced (in the majority of cases all denominators in (4.6) are in fact equal to $q_{\hat{\mu}}$, so that all coefficients become integer).

Now we iterate this process: compute all charges of the boundary vectors with respect to all basis vectors, and as soon as we encounter one with charge $1/q$, we interchange the corresponding basis vector and boundary vector. Note that in every step the determinant of the inner product matrix of the basis vectors (which is the square of the volume of the unit cell of the lattice) is reduced⁵ by a factor q^2 . This means that the procedure must end after a finite number of steps.

The only way the procedure can fail is if no charge $1/q$ can be found. A simple example demonstrating such a failure is a one-axion case with just two boundary vectors $v_1 = (2)$ and $v_2 = (3)$. There are two possible bases, and the only charges we encounter are either $\frac{2}{3}$ or $\frac{3}{2}$. This situation never occurs for any of the 32990 Gepner orientifolds. However, it may also happen that an integer basis exists, but that the algorithm converges to an incorrect basis. We did indeed encounter just three cases where we ended up with a basis with respect to which all charges are either integer, or half-integer, with values $q/2$, $|q| \geq 3$. Then no further progress is possible. These three cases could be handled by reordering the initial set of boundaries, so that the algorithm converges to a different set. For all 32990 orientifolds a maximum of 19 iterations was necessary to reach an integer basis.

Note that all charges are defined in terms of boundary vectors, as announced in (3.9), through

$$V_{a\nu} = \sum_{\mu} Q_{a\mu} V_{c(\mu)\nu} \quad (4.8)$$

so that the original basis in which the boundary vectors are expressed is irrelevant. In particular, the unusual normalization procedure of the columns drops out between the

⁵Proof: Consider the lattice spanned by the $N_A - 1$ vectors $V_{c(\mu)}$, with $V_{c(\hat{\mu})}$ removed. The volume of the full unit cell is the volume of the unit cell in this $N_A - 1$ dimensional sub-lattice, times the length of $V_{c(\hat{\mu})}$ times $\sin \theta$, where θ is the angle between $V_{c(\hat{\mu})}$ and the plane of the sub-lattice. The new vector W can be decomposed in a component along $V_{c(\hat{\mu})}$ and a component in the plane of the sub-lattice. The component of W along $V_{c(\hat{\mu})}$ has a length $1/q$ of $V_{c(\hat{\mu})}$, and the projection on the sub-lattice is irrelevant for the computation of the volume. Hence the volume decreases by $1/q$.

left- and righthand side. However, this normalization procedure lead to rational inner products, which was a great convenience for obtaining the result.

A final comment is that, although the existence of a basis is expected on general grounds (e.g. extrapolation from the geometric regime, or ultimately, from brane charge quantization), it is very remarkable that in the present setup the basis is realized in terms of RCFT boundary states. In particular, this implies that at the Gepner point all the basis vectors are mutually supersymmetric, a very special configuration reminiscent of fractional branes at singularities. It is possible that work along the lines of [67], realizing rational boundary states as fractional branes in LG orbifolds, further clarifies the nature of the above basis beyond the brute force construction.

4.1 Example: The Quintic

In order to explain how the algorithm explained above works in practice, we will consider here the quintic, a well studied case in the comparison between Gepner models and Calabi-Yau compactifications. An integral basis for tadpole charges for this case was presented in [57]. Here we need only a subset of those charges, since we are only interested in the “imaginary” boundary combinations $a - a^c$. However, the results of [57] are not sufficiently explicit to make a direct comparison possible, and furthermore there will in any case be a basis dependence.

The quintic Calabi-Yau has Hodge numbers $(h_{11}, h_{21}) = (1, 101)$, and can be obtained from the Gepner model $(3, 3, 3, 3, 3)$. As a RCFT, this has 4000 boundary states. The total number of independent axions with $R_{a(m,J)} - R_{a^c(m,J)}$ couplings turns out to 100. Of the 4000 boundary states, 32 have a Chan-Paton group $SO(N)$, and these do not couple to these axions. The remaining 3968 boundaries are pairwise related by conjugation. Hence we end up with a total of 1984 vectors $V_{a\nu}$, with $\nu = 1, \dots, 100$. We normalize them in the way explained above.

In order to have the best possible chance of finding the basis we first order the 1984 vectors in a convenient way. One would naively expect the basis vectors to have the smallest norm, so we order the 1984 $V_{a\nu}$ according to increasing norm, respecting the original CFT ordering in case of degeneracies. Then we select the first 100 independent vectors out of this set. It turns out that the first 46 are independent, and then we have to go up to number 200 to complete the set. Now we compute the charges of all 1984 vectors with respect to this basis. To do so, we start with (4.8) and contract both sides with the would-be basis vectors $V_{c(\mu)\nu}$. In this way the coefficients on both sides of the equation are related to elements of the matrix N_{ab} , eqn (4.1), which are rational. The

definition of the charges now becomes

$$N_{ac(\nu)} = \sum_{\mu} Q_{a\mu} N_{c(\mu)c(\nu)} \quad (4.9)$$

We now invert the rational matrix $N_{c(\mu)c(\nu)}$. In this case it is a 100×100 matrix, which can be inverted exactly on a computer using unlimited size integer numerators and denominators. In this way we can avoid accuracy problems with real numbers. This is essential, because in the most difficult case we have to deal with a 480×480 matrix. Using the inverse we now compute the charges. Obviously the charges of the basis vectors themselves are integers by construction, $Q_{c(\nu)\mu} = \delta_{\nu\mu}$, but this leaves 1884 non-trivial vectors to be checked, each with 100 charges. In this example, boundaries $1, \dots, 46$ are in the basis, boundary 47 is not, but turns out to have integral charges, but boundary 48 has charge $\frac{1}{2}$ with respect to the second basis vector. So following the algorithm explained above we now take boundary 48 as our second basis vector. We recompute the inverse of the new matrix $N_{c'(\mu)c'(\nu)}$, where c' denote the new basis choice. In the next iteration boundary 53 turns out to have charge $\frac{1}{2}$ with respect to basis vector 15. So we put it in the basis and try again. Now boundary 104 turns out to have charge $\frac{3}{2}$ with respect to basis vector 6, a charge that is unsuitable, but it has charge $-\frac{1}{2}$ with respect to basis vector 7. After putting boundary 104 in the basis instead of this vector, we find that all 1984 boundaries now have integer charges.

5 Results

In [48] 19345 distinct chiral classes of brane configurations were found that agree with the standard model chirally⁶. These spectra are distinguished by comparing them modulo non-chiral (vector-like) matter. They all contain a group $SU(3) \times SU(2) \times U(1)$, and all matter that is chiral with respect to that group must form *exactly* three families of quarks and charged leptons. They are distinguished by their complete Chan-Paton group, the chiral matter with respect to that group, and the massless vector bosons that exists in addition to Y . These Chan-Paton *chiral* spectra may contain matter that is $SU(3) \times SU(2) \times U(1)$ non-chiral, such as Higgs pairs and right-handed neutrinos, as well as less desirable vector-like particles. Individual models in each class differ in

⁶All spectra are available online at http://www.nikhef.nl/~t58/Site/String_Spectra.html. They were assigned a unique number to identify them, and to which we will refer henceforth. To examine an explicit sample of a spectrum in one of the 19345 classes, follow the instructions given on this webpage.

their fully *non-chiral* spectra, *i.e.* in matter that is vector-like with respect to the full, unbroken Chan-Paton gauge group. We can now in principle investigate all of them for the presence of discrete symmetries. Note that in each class there are many explicit realizations, which were collected by examining a subset of the 32990 non-zero tension orientifolds. This subset was determined by limiting, for purely practical reasons, the total number of boundary states to 1750.

The presence of discrete symmetries is not a property of the entire class, but must be examined for each class member separately. In fact, according to the search philosophy of [48] it would be natural to split these classes into subsets with definite discrete symmetries. At present, only the presence of additional $U(1)$ bosons, in other words, the constraint $\sum_a n_a N_a Q_{a\mu} = 0$, is used to distinguish classes. A natural refinement would be to distinguish brane configurations on the basis of the constraint $\sum_a n_a N_a Q_{a\mu} = 0 \bmod N$.

Since this refinement was not taken into account we have to examine classes of interest *a posteriori*. We did not do that for all 19345 classes, but limited ourselves to classes with Chan-Paton groups $U(3) \times U(2) \times U(1) \times U(1)$ or $U(3) \times Sp(2) \times U(1) \times U(1)$ (UUUU and USUU for short) with all chiral matter in bi-fundamentals, and $SU(5)$ GUT models with Chan-Paton group $U(5) \times O(1)$, $U(5) \times U(1)$, $U(5) \times U(1) \times O(1)$ or $U(5) \times U(1) \times U(1)$.

5.1 UUUU and USUU spectra

The majority of the spectra in this class are of the “Madrid” type [24]. In this configuration all U and D antiquark open strings have one end on the $U(3)$ brane and the other on one of the $U(1)$ branes (the **c**-brane, conventionally⁷), whereas the charged leptons and neutrinos have their endpoints on the **c**-brane and the **d**-brane. There is a more exotic, but also more problematic possibility of having some of the anti-quarks end on the **c**-brane and some on the **d**-brane. This class was investigated in [61]. In all but one of these models the standard model Y charge is given by

$$Y = \frac{1}{6}Y_{\mathbf{a}} - \frac{1}{2}Y_{\mathbf{c}} + \frac{1}{2}Y_{\mathbf{d}}, \quad (5.1)$$

where Y_x is the $U(1)$ generator of brane x . The signs are convention dependent, and are chosen differently in some of the literature. There is one model in the database (Nr. 13395) with a $U(3) \times U(2) \times U(1) \times U(1)$ Chan-Paton group and an unconventional Y

⁷We use boldface subscripts to refer to one of the four standard model branes.

charge $Y = -\frac{1}{3}Y_{\mathbf{a}} + \frac{1}{2}Y_{\mathbf{b}} + Y_{\mathbf{d}}$. There are just eight samples of this particular model, and none of them had discrete symmetries, so we will not consider it here.

In table 1 we summarize all 24 models with the hypercharge embedding chosen as in (5.1). The horizontal lines separate the USUU and the UUUU models, and within these sets they separate the standard Madrid models (with perturbative lepton number conservation) from the non-standard ones. In the columns labelled \mathbf{xy} etc. we list the chiral intersection number $I_{\mathbf{xy}}$. If this number is negative, this implies that both endpoint branes must be conjugated. Note that in addition there may always be non-chiral matter, which is ignored in the definition of a chiral class. Within the USUU class, there is only one Madrid model possible, but in the UUUU class there are several, depending on the choice of the $U(2)$ representation (2) or (2*) used for the standard model matter. Furthermore the Higgs bosons can be chiral with respect to $U(2)_{\mathbf{b}}$, and hence contribute to $U(2)$ anomaly cancellation. No restriction was imposed on the number of such “chiral” Higgs bosons. Even if the number is zero, Higgs bosons can still occur as $U(2)$ -non-chiral particles in the spectrum.

For each model, there is a possibility of having just a single massless $U(1)$ boson, Y , or two, Y and $B-L$. The latter possibility is much more common than the second, and for some configurations the first option was not realized at all in the set of Gepner orientifolds explored in [48]. In models that are not strictly of the Madrid type there is usually just a single non-anomalous $U(1)$ and hence only Y is gauged. In models 7488 and 13015 however, there is an additional anomaly free $U(1)$, namely $Q_{\mathbf{b}} + 2Q_{\mathbf{c}}$, which is unbroken in the second class. Interestingly, the unbroken case occurred less often than the broken case, just the other way around as for Madrid models⁸. There are several conventions one can choose to represent these spectra. One may conjugate the \mathbf{b} brane, which does not participate in Y . Furthermore, one can conjugate the \mathbf{d} brane and simultaneously interchange the assignment of right-handed neutrinos and charged leptons, thus keeping Y unchanged. The same can be done with the \mathbf{c} brane. Our conventions are such that in the multiplicities of the fields are as much as possible positive. In comparison with [22], formula (3.14), the first option corresponds to numbers 10551 and 1352, after conjugation of \mathbf{d} . The second option, shown in parentheses in [22], corresponds to 12106 and 7976.

⁸Since a massless $U(1)$ requires the set of axion charges to have a null vector, one would generally expect spectra with fewer massless $U(1)$ ’s to occur more frequently. Therefore Madrid models are probably the exception and not the rule. Presumably this is due to the fact that in Madrid models the \mathbf{a} and \mathbf{d} play a symmetric role, so that often their axion couplings are identical. This implies a massless $B-L$. In particular this is true for a large subclass related to Pati-Salam models.

Note that in non-Madrid models lepton number is not defined in terms of brane charges, and hence there is an intrinsic confusion between L , H_d and the conjugate of H_u . These spectra are accepted as standard-model-like on the basis of a correct count of the net number of $(1, 2, -\frac{1}{2})$ representations, *i.e.* $I_{\mathbf{bc}^*} + I_{\mathbf{b}^*\mathbf{c}^*} + I_{\mathbf{bd}^*} + I_{\mathbf{b}^*\mathbf{d}^*} = 3$. Exactly which particles should be identified as Higgses or lepton doublets depends on the superpotential couplings and on the direction of the Higgs vev in the space of these fields. This requires additional assumptions. In [61] this was discussed for some models in class 14062.

How generic are discrete symmetries in these 24 models? The total number of 24 classes splits into 18 with a unitary weak group and 6 with a symplectic one. In total, these classes contain 962958 brane configurations. The discrete symmetries we have encountered are \mathbb{Z}_2 and \mathbb{Z}_3 and occur only for 6 of these 24 models. We only consider a discrete symmetry if it is not contained in the continuous symmetries. With Y of the form (5.1), there is automatically a null-vector n^a of the form $(1, 0, -3, 3)$. Reduced modulo 3 this yields $(1, 0, 0, 0)$, and hence there is automatically a \mathbb{Z}_3 symmetry corresponding to $n^a = (1, 0, 0, 0)$. This is of no interest, since it is just the $SU(3)$ color selection rule requiring that all amplitudes be color (and hence triality) singlets. Not surprisingly, this also follows from Y -charge conservation. Similarly, in models with a $U(2)$ factor there is a \mathbb{Z}_2 null vector $n^a = (0, 1, 0, 0)$. This does not follow from any continuous charge conservation, but it is equally uninteresting, since it just imposes $SU(2)$ duality. Both of these discrete symmetries are a direct consequence of the factors N_a in (3.10). We do not include them in our count. Apart from these, the total number of \mathbb{Z}_2 symmetries we find is 2152 (0.2% of the total), and the total number of \mathbb{Z}_3 symmetries 61664 (6.4% of the total).

The \mathbb{Z}_2 only occurred in the class $U(3) \times Sp(2) \times U(1) \times U(1)$ with a massive $B-L$ boson, as a subgroup of $B-L$. This is just conventional R -parity. However, the complete set is dominated by models with a massless $B-L$ boson, class 2751. Here the standard \mathbb{Z}_2 R -parity symmetry is already contained in $B-L$. If we were to exclude all classes with unbroken $B-L$, there are just 46990 left. In this set, about 4.6% of the models have a non-trivial \mathbb{Z}_2 discrete symmetry (not including $U(2)_{\mathbf{b}}$ duality). But even then, we must conclude that in these models discrete symmetries are a fairly rare phenomenon, occurring in only a few percent of the cases. The other conclusion is that \mathbb{Z}_3 discrete symmetries are about as common as \mathbb{Z}_2 discrete symmetries, at least in this region in the landscape. In table 1 we list for all 24 classes how many samples there are in the database of [48], and how many of these have discrete symmetries.

We will now describe each of these cases in a bit more detail. Note that the numbers in table 1 specify the number of distinct brane label combinations (**a, b, c, d**) that yield a given chiral spectrum. Those spectra are in principle non-chirally distinct (although in practice there are often huge degeneracies), and also the precise axion couplings may be different. But the differences are small, and hence it suffices to present just one example per class. In order to make the results reproducible, we specify for each example the tensor product, MIPF and orientifold, and the brane labels for which they occur. However, it is difficult to present this information in a basis-independent way. Instead we give labels as used and recognized by the computer program `kac` used to produce these spectra, and which is publicly available. Instructions for exactly reproducing these spectra can be found on the webpage http://www.nikhef.nl/~t58/Site/String_Spectra.html.

5.1.1 Examples: USUU and UUUU models

We now turn to several illustrative examples, and their discrete symmetries, which are classified according to the notation in [2]. To help identify these examples we specify the Hodge numbers of the corresponding Calabi-Yau manifold, by comparing the closed string spectrum to a type-IIA compactification. In addition, we specify how the h_{11} $N = 2$ hyper multiplets split into chiral multiplets and vector multiplets. Precise definitions of all these quantities in terms of the partition function are given in the appendix. In each case we indicate which couplings of phenomenological interest are perturbatively allowed, which ones are forbidden by the discrete symmetry, and which ones are non-perturbatively allowed. Couplings in the latter category are not forbidden by any discrete symmetry, and hence can in principle be generated by instantons. However, we are not claiming that those instantons actually exist in a given model.

Example 1: \mathbb{Z}_2 in $U(3) \times Sp(2) \times U(1) \times U(1)$ with broken $B-L$ (class 7506)

An example was found for tensor product 241446, MIPF 10, Orientifold 2, boundary states (630, 41, 1070, 631). The Hodge numbers of the corresponding Calabi-Yau manifold are $h_{21} = 28$, $h_{11} = 40$, and in the orientifold $h_{11}^+ = 35$ (leading to 35 Kahler moduli) and $h_{11}^- = 5$ (5 RR vector bosons). In this class all Yukawa couplings are perturbatively allowed, as is the μ -term. This is generally true in $USUU$ -type Madrid models. The \mathbb{Z}_2 is a subgroup of the broken $B-L$, and this is standard R-parity. All dimension-4 baryon and lepton violating couplings are forbidden, including LH_u , but the couplings $QQQL$ and $UUDE$ are non-perturbatively allowed. On the other hand

Majorana neutrino masses, as well as the Weinberg operator LLH_uH_u which can also give rise to such masses is allowed. All odd powers of the neutrino superfield are forbidden. In this case there are 16 independent axions (out of a total of 28+1) coupling non-trivially to branes, and the couplings to the standard model branes is as follows.

$$\begin{array}{ll}
\mathbf{a:} & 0 \quad -3 \quad 0 \quad 0 \quad -3 \quad -3 \quad -3 \quad 3 \quad 0 \quad -3 \quad 0 \quad 3 \quad 3 \quad 6 \quad 3 \quad 6 \\
\mathbf{c:} & 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad -2 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \\
\mathbf{d:} & 0 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \quad -1 \quad -1 \quad 0 \quad 1 \quad 0 \quad -1 \quad -1 \quad -2 \quad -1 \quad -2
\end{array} \tag{5.2}$$

Example 2: \mathbb{Z}_3 in $U(3) \times Sp(2) \times U(1) \times U(1)$ with broken $B-L$ (class 7506)

The \mathbb{Z}_3 discrete symmetries in this class do not overlap with the \mathbb{Z}_2 symmetries described above. An example with \mathbb{Z}_3 symmetries is tensor 2101010; MIPF 63; Orientifold 0, boundaries (192, 503, 227, 237) and Hodge numbers $h_{21} = 7$, $h_{11} = 67$, $h_{11}^+ = 64$ and $h_{11}^- = 3$. The \mathbb{Z}_3 nul vector is (0,1,1), as one can read off from the axion couplings.

$$\begin{array}{ll}
\mathbf{a:} & 0 \quad 0 \quad -6 \quad 0 \quad 3 \\
\mathbf{c:} & -6 \quad 6 \quad 5 \quad -3 \quad -4 \\
\mathbf{d:} & -6 \quad 6 \quad 7 \quad -3 \quad -5
\end{array} \tag{5.3}$$

This symmetry corresponds to $R_3L_3^2$ in table 2 of [22]. Note that the definition of the generator L in this paper differs by a sign from the standard definition of lepton number, which might easily lead to a confusion between R_3L_3 and $R_3L_3^2$. This has been taken into account, and furthermore we have checked explicitly that the discrete symmetries forbid all couplings of type UDD , QDL , LLE , LH_u , $QQQL$, and $UUDE$, confirming that it indeed corresponds to $R_3L_3^2$. The μ term and all Yukawas are perturbatively allowed but neutrino Majorana masses and the Weinberg operator are forbidden. Up to fourth order, there is just one coupling that is perturbatively forbidden but non-perturbatively allowed, and that is the third power of the neutrino superfield.

Example 3: \mathbb{Z}_3 in $U(3) \times Sp(2) \times U(1) \times U(1)$ with unbroken $B-L$ (class 2751)

This example occurs for tensor 2101010; MIPF 64 and orientifold 0 and boundaries (46, 5, 48, 415), Hodge numbers $h_{21} = 7$, $h_{11} = 43$, $h_{11}^+ = 40$ and $h_{11}^- = 3$. All $B-L$ violating couplings are forbidden by the unbroken $B-L$, and the operators $QQQL$ and

$UUDE$ are forbidden by the discrete symmetry. All Majorana neutrino mass contributions are also forbidden by $B-L$. The μ -term and all Yukawas are perturbatively allowed. The axion couplings in this example are:

$$\begin{array}{rcl}
\mathbf{a}: & 9 & 0 & 0 & 0 & 0 \\
\mathbf{c}: & 0 & 0 & 0 & 0 & 0 \\
\mathbf{d}: & 3 & 0 & 0 & 0 & 0
\end{array} \tag{5.4}$$

This obviously has an L_3 discrete symmetry: lepton number can only be violated in units of three. Since $B-L$ is conserved the same is then automatically true for baryon number as well. Note that Y conservation forces the \mathbf{a} -brane couplings to be three times those of the \mathbf{d} . The resulting factor 9 incorporates both $SU(3)$ triality and the B_3 discrete symmetry.

Example 4: \mathbb{Z}_3 in $U(3) \times U(2) \times U(1) \times U(1)$ with unbroken $B-L$ (class 1352)

An example of this kind occurs for tensor 2101010; MIPF 59; Orientifold 0, boundaries (932, 650, 881, 1302). The Hodge numbers are $h_{21} = 19$, $h_{31} = 59$, $h_{11}^+ = 29$ and $h_{11}^- = 2$. The axion couplings are:

$$\begin{array}{rcl}
\mathbf{a}: & 0 & 0 & 0 & 9 & 0 & 0 & 0 & 9 & 0 & 0 & 9 & 0 & 0 & 0 & 9 & 9 & 9 \\
\mathbf{b}: & 2 & 2 & 2 & 4 & 4 & 2 & 0 & 0 & 4 & 0 & 2 & 0 & 2 & 2 & 0 & 0 & 0 \\
\mathbf{c}: & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{d}: & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 3 & 0 & 0 & 3 & 0 & 0 & 0 & 3 & 3 & 3
\end{array} \tag{5.5}$$

from which we can read off that there is indeed a massless Y and $B-L$, and that furthermore there is a \mathbb{Z}_2 null vector $(0, 1, 0, 0)$ and a \mathbb{Z}_3 null vector $(0, 0, 0, 1)$. The former just imposes $SU(2)$ duality, and the second corresponds to an L_3 discrete symmetry, as in example 3. In this class of models, two of the three quark masses (for up as well as down quarks) must be generated non-perturbatively, but the discrete symmetries do not forbid that. All lepton Yukawas are perturbatively allowed. We are assuming here, as in the USUU examples, that the Higgs comes from the non-chiral spectrum, from bi-fundamentals between the \mathbf{b} and \mathbf{c} branes. The $B-L$ forbids the usual dimension-4 terms as well as Majorana neutrino masses, while $QQQL$ and $UUDE$ are forbidden by the \mathbb{Z}_3 symmetry. A μ -term can in principle be generated non-perturbatively, and is not forbidden by the \mathbb{Z}_3 .

Example 5: \mathbb{Z}_3 in $U(3) \times U(2) \times U(1) \times U(1)$ with unbroken $B-L$ (class 7976)

This example occurs for tensor 242222; MIPF 16; orientifold 2 and boundaries (343, 6, 610, 436), Hodge numbers $h_{21} = 21, h_{11} = 69, h_{11}^+ = 52$ and $h_{11}^- = 17$. The axion couplings are

$$\begin{array}{ll}
\mathbf{a}: & 0 \ 0 \ 18 \ 0 \ 0 \ 0 \ 0 \ 9 \ 9 \ -18 \ 0 \ 0 \ -9 \ 9 \ 9 \\
\mathbf{b}: & 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ -2 \ 0 \\
\mathbf{c}: & 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \\
\mathbf{d}: & 0 \ 0 \ -6 \ 0 \ 0 \ 0 \ 0 \ -3 \ -3 \ 6 \ 0 \ 0 \ 3 \ -3 \ -3
\end{array} \tag{5.6}$$

The rest of the discussion is similar for example 4. There is a massless $B-L$ forbidding the usual couplings, as above. However, in this case the Higgs pair comes out automatically within the chiral spectrum. Perturbative Yukawa couplings appear for two of the up and down-quarks and one of the leptons, and the missing ones are in principle allowed non-perturbatively. Also a μ -term may be generated non-perturbatively. However all $QQQL$ and $UUDE$ terms are forbidden by the discrete symmetries.

Example 6: \mathbb{Z}_3 in $U(3) \times U(2) \times U(1) \times U(1)$ with unbroken $B-L$ (class 14792)

The next example also has a \mathbb{Z}_3 symmetry, but it is a bit different from the foregoing two. It occurs for tensor 242222; MIPF 16; orientifold 2, boundaries (284, 343, 700, 335), Hodge numbers $h_{21} = 21, h_{11} = 69, h_{11}^+ = 52$ and $h_{17}^- = 2$ (as in the previous example).

$$\begin{array}{ll}
\mathbf{a}: & -3 \ -3 \ 27 \ 3 \ 9 \ -6 \ -6 \ 18 \ 21 \ -33 \ 0 \ 6 \ -9 \ 0 \ 3 \\
\mathbf{b}: & 0 \ 0 \ 12 \ 0 \ 0 \ 0 \ 0 \ 6 \ 6 \ -12 \ 0 \ 0 \ -6 \ 6 \ 6 \\
\mathbf{c}: & 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \\
\mathbf{d}: & -1 \ -1 \ 9 \ 1 \ 3 \ -2 \ -2 \ 6 \ 7 \ -11 \ 0 \ 2 \ -3 \ 0 \ 1
\end{array} \tag{5.7}$$

In this example, the \mathbf{b} -brane couplings have a \mathbb{Z}_6 symmetry. The \mathbb{Z}_2 subgroup is the usual uninteresting $SU(2)$ duality. However, the \mathbb{Z}_6 symmetry forbids any operator built out of two or four $U(2)$ doublets with the same $U(1)$ charge, such as $QQQL$. On the other hand, it does not forbid the operator $UUDE$. The dimension-4 operators are forbidden by the unbroken $B-L$. This discrete symmetry is not explicitly mentioned in [22] because these authors allowed at most one $U(2)$ -chiral Higgs pair, and this model has six. This discrete \mathbb{Z}_6 symmetry is anomaly free as long as the number of families is a multiple of three, and the number of Higgs doublets a multiple of six, which is the case here (if it were not the case a \mathbb{Z}_6 symmetry could not have appeared). We assume that one of these six Higgs candidates plays the rôle of the Higgs boson. Then

all quark and lepton Yukawa couplings are perturbatively allowed. A μ -term however cannot be generated even non-perturbatively because of the discrete symmetry.

Example 7: \mathbb{Z}_3 in $U(3) \times U(2) \times U(1) \times U(1)$ (exotic) (class 7488)

This is an example of discrete symmetries occurring for a non-Madrid model. It was found for tensor 441010; MIPF 21; Orientifold 1, Hodge numbers $h_{21} = 28, h_{11} = 10, h_{11}^+ = 10$ and $h_{11}^- = 0$

$$\begin{array}{rcl}
\mathbf{a:} & -3 & -6 & -3 & 3 & 0 & 0 \\
\mathbf{b:} & -4 & -2 & 6 & 0 & -4 & 2 \\
\mathbf{c:} & 1 & -1 & -1 & 1 & -1 & -1 \\
\mathbf{d:} & 2 & 1 & 0 & 0 & -1 & -1
\end{array} \tag{5.8}$$

The only massless $U(1)$ is Y . Furthermore, we find the usual $SU(2)$ -duality \mathbb{Z}_2 , but also a \mathbb{Z}_3 null vector is $(0, 1, 0, 2)$. Note that the \mathbf{d} -brane cannot be associated with lepton number, since this is not a Madrid model. Furthermore there is no family universality for the up and down quarks. This has the consequence that there is no universal rule for the dimension 4 couplings: some of the UDD couplings are non-perturbatively allowed, and some are forbidden by discrete symmetries; some of the QDL couplings are perturbatively allowed, others are non-perturbatively allowed, and some are forbidden by the discrete \mathbb{Z}_3 symmetry. All LLE coupling are non-perturbatively allowed. Some of the $QQQL$ and $UUDE$ couplings are non-perturbatively allowed, and some are forbidden by the \mathbb{Z}_3 symmetry. We cannot discuss Higgs couplings in general, because none of the particles in the Chan-Paton chiral spectra can play the rôle of the Higgs. The three particles in the \mathbf{bc} intersection must be interpreted as lepton doublets. Clearly the Higgs must come from the non-chiral spectrum, and could come from any strings ending on the \mathbf{b} brane en with the other end on the \mathbf{c} or \mathbf{d} brane. Since lepton number is undefined, there is no obvious way of deciding this. The precise phenomenological fate of this type of model is hard to assess, because of the family-dependent presence of baryon and lepton number violating couplings, and the fact that there is no obvious Higgs candidate.

5.2 GUT spectra

Orientifold spectra with an $SU(5)$ GUT spectrum have been studied in many papers [68, 69, 70, 71, 50, 51]. In [22] the possible presence of discrete symmetries in models was discussed. There are two important issues here, one in which discrete symmetries

would be catastrophic, and another where they would be beneficial. The first concerns the up-quark Yukawa couplings⁹ which in these models are forbidden perturbatively, but may be generated by instantons [30, 33]. A discrete symmetry might forbid the existence of these instantons. On the other hand, discrete symmetries to forbid the usual dimension four B and/or L violating couplings would be very welcome. For $SU(5)$ GUT models baryon number violating MSSM couplings are *a priori* even more threatening, because the same instantons that generate the Yukawa couplings also tend to generate the $QQQL$ term [50]. Here discrete symmetries are less obviously useful, because they would tend to forbid the up-quark Yukawas as well. However, in [51] it was pointed out that specially chose brane realizations can help solving this problem.

Here we examine the presence of discrete symmetries for certain Gepner model GUT realizations, namely all the two-stack models where the second stack is either $U(1)$ or $O(1)$, and a subset of the three-stack models.

5.2.1 Two-stack models

The simplest realizations of $SU(5)$ GUT consist of one stack of five branes producing a $U(5)$ Chan-Paton group, and a second stack which does not couple to any of the standard model gauge interactions. The chiral matter consists of three $(\mathbf{10})$'s of $SU(5)$, plus three bi-fundamentals from open strings stretched between the two stacks that are in the $(\mathbf{5}^*)$ of $SU(5)$. In addition there may be a number of $(\mathbf{5}) + (\mathbf{5}^*)$ pairs that play the rôle of Higgs bosons. In the database of [48] there are five distinct types of such models with an $U(1)$ or $O(1)$ second stack. These were discussed in some detail in [50]. There also exist two-stack models where the second stack is $U(3)$ or $O(3)$, with the multiplicity of the $(\mathbf{5}^*)$ originating from the extra brane group, but this leads to additional complications, and we will not consider them here. The chiral spectrum of the five classes is shown in table (2).

We have examined this entire class for the presence of discrete symmetries, and found only a few examples for one of the $U(5) \times U(1)$ classes, namely nr. 345 (In [50] the models were numbered by frequency, and this class was referred to as nr. 2753, with a total of 1136 samples in the database). The discrete symmetry is a \mathbb{Z}_2 embedded in the two $U(1)$ factors. However, the chiral matter is in antisymmetric tensors of $U(5)$, bi-fundamentals and symmetric tensors of $U(1)$ which are all uncharged with respect to this \mathbb{Z}_2 . Hence there is no chance of forbidding any couplings. This implies on

⁹In flipped $SU(5)$ models these remarks apply to the down-quark Yukawa couplings

the one hand that there is no obstacle to generating the perturbatively forbidden up-quark Yukawa couplings, but on the other hand that there is no chance of forbidding dangerous B and/or L violating operators. Examples of instantons generating these Yukawas were indeed found in [50], but only in six case tadpole canceling hidden sectors could be added, and only by allowing chiral observable-hidden matter (which becomes vector-like when reduced to the standard model).

5.2.2 Three-stack models

If we allow two additional stacks instead of one the number of possibilities becomes much larger. There are 257 classes with a Chan-Paton group $U(5) \times U(1) \times U(1)$ and 168 with a group $U(5) \times U(1) \times O(1)$. If we allow higher multiplicities for the extra branes there are even more possibilities. Here we will only consider the subclass studied in [51], in which the up-quark Yukawa couplings can be generated by instantons without generating $QQQL$ couplings. This class consists of seven chiral types, listed in table 2. All models in the class considered here have natural Higgs candidates in the chiral spectrum coming from $\mathbf{a}^*\mathbf{b}^*$ and \mathbf{ac} bi-fundamentals. There are always three Higgs pairs. Here we chose to assign the $\mathbf{a}^*\mathbf{b}$ bi-fundamentals to the $(\mathbf{5}^*)$ of $SU(5)$ containing the lepton doublet and the anti-down quarks. Note that there are always three mirror pairs of particles with the quantum numbers of down quarks which form an $SU(5)$ multiplet together with the three Higgs pairs. This is the usual doublet-triplet splitting problem. In analyzing couplings, we have to make sure to take the down quarks from the $\mathbf{a}^*\mathbf{b}$ bi-fundamentals, and assume that the others pair off into massive particles. Matter from the \mathbf{bc} sector is neutral, and could play the rôle of neutrinos, although there are either six or eight such states. In addition, in model 4325 there is a symmetric tensor, which is another neutrino candidate. Note that the anti-symmetric tensors on the \mathbf{b} and \mathbf{c} branes listed in table 2 correspond to string sectors without massless states. They are merely listed here because there do exist massive states in these sectors, and because they contribute to anomaly cancellation.

5.2.3 Examples: GUT models

Here we present for each distinct class of discrete symmetries one example in some detail. The model numbers (“class”) refer to table 2. Note that the discrete symmetries in the only two-stack example in table 2, model 345, do not forbid anything, as explained above. So we do not present that example in more detail.

Example 8: \mathbb{Z}_2 in $U(5) \times U(1) \times O(1)$ (class 57)

This was found for tensor 1102222; MIPF 18; orientifold 0 with boundaries numbers (566, 566, 1308, 990). The Hodge numbers are $h_{21} = 13, h_{11} = 109, h_{11}^+ = 98$ and $h_{11}^- = 11$. Note that the first two boundary labels are here for the standard model $U(3)$ and $U(2)$ stacks. The fact that they are identical implies that we have $U(5)$. In the following examples we will always combine these two groups to $U(5)$ when writing axion couplings. Then **a** denotes the $U(5)$ stack, **b** the first $U(1)$, and **c** the second $U(1)$ (if any). The axion couplings are

$$\begin{array}{lcl} \mathbf{a}: & 5 & 0 & -10 & 0 & 5 & -10 & -5 & 5 & -20 & 0 \\ \mathbf{b}: & 0 & -2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \quad (5.9)$$

There is a surviving \mathbb{Z}_2 symmetry associated with the $U(1)$ brane. This symmetry forbids the LH_u term, the μ -term, as well as all $QQQL$ and $UUDE$ terms. Down quark and lepton Yukawa couplings are perturbatively allowed, and up-quark Yukawas are non-perturbatively allowed. But on the other hand, UDD , QDL and LLE terms are non-perturbatively allowed as well, although they are generated by different instantons then the Yukawa couplings, and hence could have a different strength. Since the μ term is perturbatively forbidden, it follows from $SU(5)$ symmetry that also a mass term for the color triplet partners of the Higgses is forbidden. So here the discrete symmetry has a negative effect. In the down quark sector we get $6(d^*) + 3(d)$, and all options for pairing off the vector-like d -quarks are forbidden by the discrete symmetry.

Example 9: \mathbb{Z}_3 in $U(5) \times U(1) \times U(1)$ (class 4004)

This was found for tensor 1102222; MIPF 28, orientifold 0 and boundaries numbers (816, 816, 1309, 917). The Hodge numbers are $h_{21} = 13, h_{11} = 37, h_{11}^+ = 34$ and $h_{11}^- = 3$. The axion couplings are

$$\begin{array}{lcl} \mathbf{a}: & 0 & 5 & -10 & -5 & 0 & 5 & 5 & -10 & 0 & -5 \\ \mathbf{b}: & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \mathbf{c}: & 0 & 1 & 1 & -1 & 0 & 1 & 1 & -2 & 0 & -1 \end{array} \quad (5.10)$$

Note that the vector boson coupling to Y_b remains massless. This forbids many superpotential terms. There is a surviving \mathbb{Z}_3 symmetry embedded in the combination of the $U(1)$ -generator of the **a** stack and the one of the **c** stack. This symmetry forbids masses for the neutrino candidates from the **b-c** intersection. Majorana masses for these six states are already forbidden by the masses Y_b , but the discrete symmetry also prevents them from pairing off into a Dirac mass term. Nothing else of any interest is forbidden

by the discrete symmetry. Yukawa couplings are perturbatively or non-perturbatively allowed, as in the previous case.

Example 10: \mathbb{Z}_2 in $U(5) \times U(1) \times U(1)$ (class 4316)

It was found for tensor 1102222; MIPF 18, orientifold 0, and boundary numbers (523, 523, 1307, 566), Hodge numbers $h_{21} = 13, h_{11} = 109, h_{11}^+ = 98$ and $h_{11}^- = 11$. The axion couplings are:

$$\begin{array}{ll} \mathbf{a}: & 5 \quad 0 \quad -10 \quad 0 \quad 5 \quad -10 \quad -5 \quad 5 \quad -20 \quad 0 \\ \mathbf{b}: & 0 \quad 2 \quad 2 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \\ \mathbf{c}: & 1 \quad 0 \quad -2 \quad 0 \quad 1 \quad -2 \quad -1 \quad 1 \quad -4 \quad 0 \end{array} \quad (5.11)$$

This case is similar to the foregoing one, except that $Y_{\mathbf{a}} - 5Y_{\mathbf{c}}$ rather than $Y_{\mathbf{b}}$ remains massless. There is a surviving \mathbb{Z}_2 symmetry in $Y_{\mathbf{b}}$, but no couplings are forbidden by it that are not already forbidden by the extra $U(1)$. Interestingly, *all* fourth order superpotential terms are absent in this class of models, as well as all first and second order terms. The absence of second order terms implies in particular that mass terms for the vector-like down quarks from the Higgs multiplets are forbidden. These features are a property of the entire class, irrespective of discrete symmetries. The problem is that there are (5^*) 's from $\mathbf{a}^*\mathbf{b}$ and $\mathbf{a}^*\mathbf{b}^*$ and a (5) from \mathbf{ac} . The first two have charge -1 w.r.t. $Y_{\mathbf{a}} - 5Y_{\mathbf{c}}$, but the latter has charge -4 . This forbids any Dirac mass term. Therefore to make this kind of spectrum viable we first have to break the additional $U(1)$ symmetry, either by a Higgs mechanism, or by axion mixing directly in string theory. There are indeed example in the database where the latter occurs. This is class 4324, which will be discussed below.

Example 11: \mathbb{Z}_3 in $U(5) \times U(1) \times U(1)$ (class 4316)

This example was found for tensor 441010; MIPF 71; orientifold 0 and boundary states (485, 485, 525, 581), Hodge numbers $h_{21} = 20, h_{11} = 14, h_{11}^+ = 14$ and $h_{11}^- = 0$. The axion couplings are

$$\begin{array}{ll} \mathbf{a}: & -5 \quad -10 \quad -5 \quad -10 \quad 5 \quad 0 \quad 5 \quad 20 \quad -10 \quad -10 \\ \mathbf{b}: & 0 \quad 0 \quad -3 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \\ \mathbf{c}: & -1 \quad -2 \quad -1 \quad -2 \quad 1 \quad 0 \quad 1 \quad 4 \quad -2 \quad -2 \end{array} \quad (5.12)$$

This is similar to the previous example, except that the discrete symmetry that is embedded in $U(1)_{\mathbf{b}}$ is \mathbb{Z}_3 . As before, the superpotential only contains terms of order three, or five and higher, as a consequence of the extra $U(1)$. Hence we have the same problem with lifting the vector-like down quark pair. The discrete symmetry forbids

the terms UDD , QDL and LLE , provided that for D we use the field that is *not* in the Higgs multiplet. There are even perturbatively allowed baryon number violating couplings involving the vector-like down quark pair. Up-quark Yukawa couplings are non-perturbatively allowed, down quark Yukawas are perturbatively allowed, as are charged lepton Yukawas, and neutrino Yukawa couplings are forbidden by the discrete symmetry.

Example 12: \mathbb{Z}_2 in $U(5) \times U(1) \times U(1)$ (class 4324)

This was found for tensor 1102222; MIPF 27; Orientifold 0 and boundary states (365, 365, 1393, 572) with Hodge numbers $h_{21} = 12$, $h_{11} = 96$, $h_{11}^+ = 90$ and $h_{11}^- = 6$. This class is similar to 4316 discussed above, except that there are no extra $U(1)$ gauge bosons, and consequently many more couplings are allowed. In particular this includes the vector-like down-quark pair. Since in model 4316 this mass was forbidden by the extra $U(1)$, it follows that in class 4324 it may in principle be generated by instantons. However, that cannot happen in the eight cases with extra discrete symmetries we are discussing here. In this particular example the axion couplings are:

$$\begin{array}{ll}
\mathbf{a}: & 0 \quad 5 \quad -5 \quad -10 \quad -5 \quad 0 \quad -10 \quad 0 \quad 5 \\
\mathbf{b}: & 2 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \\
\mathbf{c}: & 0 \quad 0 \quad -1 \quad -2 \quad -1 \quad 0 \quad -2 \quad 0 \quad 1
\end{array} \tag{5.13}$$

From which we read off a discrete symmetry \mathbb{Z}_2 associated with brane **b**. This symmetry forbids the mass terms needed to lift the vector-like down-quark. It also forbids a μ -term, and all $QQQL$ and $UUDE$ terms. The up-quark Yukawas are non-perturbatively allowed, and the down quark and charged lepton Yukawas are perturbatively allowed. Note that all three-stack models we consider here satisfy a criterium discussed in [51], namely that up-quark Yukawas can be generated without automatically generating $QQQL$ and $UUDE$ term of similar strength. This is generically a problem in $U(5)$ orientifold models [50]. In the three-stack classes discussed here this problem can be solved because the up-quark Yukawas and the $QQQL$ and $UUDE$ terms are generated by instantons with different charges. Hence it is at least possible in principle that they contribute with different strengths. Here we see an even better solution to this particular problem. Precisely because these terms violate a different set of charges, it is possible for discrete symmetries to forbid one and not the other. That is exactly what is happening here. So we see here an example where discrete symmetries are playing a very useful rôle, but this is overshadowed by at least two serious problems. The first is that there are *no* discrete symmetries forbidding the UDD , QDL and LLE terms.

These are perturbatively forbidden, but may be generated by instantons. The second is the forbidden vector-like down-quark mass (*i.e.* the down quark triplets in the Higgs multiplets). The brane charge violation of the latter terms is precisely the sum of the charge violations of the up-quark Yukawas and the $QQQL$ or $UUDE$ terms. Hence any discrete symmetries that forbid the latter but not the former will automatically forbid the lifting of the down quark mirror pair. This is just a manifestation of the doublet-triplet splitting problem in $SU(5)$ models.

Example 13: \mathbb{Z}_2 in $U(5) \times U(1) \times U(1)$ (class 4325)

This was found for tensor 1102222; MIPF 27; Orientifold 0 and boundary states (365, 365, 1506, 818), with Hodge numbers $h_{21} = 12$, $h_{11} = 96$, $h_{11}^+ = 90$ and $h_{11}^- = 6$. This class is very similar to 4324. It has some additional neutral chiral matter, but as in class 4324 there is no additional massless $U(1)$. The axion couplings are

$$\begin{array}{ll} \mathbf{a}: & 0 \quad 5 \quad -5 \quad -10 \quad -5 \quad 0 \quad -10 \quad 0 \quad 5 \\ \mathbf{b}: & 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \\ \mathbf{c}: & -1 \quad 1 \quad -3 \quad -2 \quad -1 \quad 0 \quad -2 \quad 0 \quad 1 \end{array} \quad (5.14)$$

The discrete symmetry is \mathbb{Z}_2 . The corresponding null vector is $(1, 1, 1)$. This implies that all matter is uncharged with respect to it, because all matter is either in rank two tensors or bi-fundamentals. Hence no couplings are affected by this symmetry. One point worth noting is that if there are hidden sectors, any observable-hidden matter is necessarily odd under the \mathbb{Z}_2 . Hence this symmetry is like an exotic (*i.e.* observable-hidden) matter parity.¹⁰ All exotic matter can only be created in pairs, and there will be a lightest exotic state that cannot decay into standard model particles, and hence, if neutral, could be a dark matter candidate. Note that in the $U(5)$ class observable-hidden matter has integral electric charge (whereas in Madrid type models they have half-integer charge), and hence this conserved exotic matter parity is not a trivial consequence of charge conservation, nor is it in disagreement with the fact that no fractional electric charge has ever been observed. This mechanism could in principle work equally well in non-supersymmetric models (provided examples can be found) and hence this provides an alternative to the rôle of R-parity in solving the dark matter problem. Furthermore the general category to which this model belongs (the “ $x = 0$ ” category of [48]) includes plenty of examples where instead of a $U(5)$ stack there are separate $U(3)$ and $U(2)$ stacks, so that there is no $SU(5)$ relation among the couplings

¹⁰The same remark applies to the \mathbb{Z}_2 symmetry we found in the two-stack models in class 345, but which was not presented in detail.

(which would be a problem without low energy supersymmetry). However, unlike R-parity, the existence of this exotic matter parity is lacking a convincing motivation.

5.3 Tadpole Cancellation

All models we have considered so far are brane configurations, which in most cases have uncanceled tadpoles. For all cases where non-trivial discrete symmetries were found (*i.e.* those listed in the last columns of table 1 and 2) we have attempted to find hidden sectors that cancel all tadpoles, allowing at most four additional stacks. Furthermore we have allowed massless matter in the observable-hidden sector, provided that it is non-chiral with respect to the full Chan-Paton gauge group, so that in principle it can acquire a mass by moving into moduli space, without breaking any gauge symmetries. These are essentially the same criteria used in [45] and [48], except that in those searches the number of additional branes was only limited by practical considerations.¹¹ Among the non-Madrid models, only classes 14062 and 7488 contain at least one model allowing a tadpole-cancelling hidden sector.

Based on previous experiences with solving tadpole conditions in RCFT orientifolds, we expected only a relatively small success rate, around one percent or even less. Surprisingly however, the overall success rate for the models in table 1 was much higher, around 65%. We found a total of 41456 cases with tadpole-cancelling hidden sectors, for a total of 63728 configurations with discrete symmetries. On the other hand, in the class of $SU(5)$ models, we did not find a single case with tadpole cancellation and discrete symmetries.

One might be tempted to conclude, although discrete symmetries are rare within the set of standard model configurations, they are more common in the physically relevant class of fully consistent open string models. There is indeed a reason why that could be true. Both tadpole cancellation and discrete symmetry conditions are more easily satisfied if there are fewer Ishibashi states reps. axions, which correlates with smaller values of h_{12} . Hence it is quite easily imaginable that after imposing the requirement of having discrete symmetries, we are left with precisely those cases where the tadpole conditions are more easy to satisfy as well.

On the other hand, it turns out that especially in the set of 59808 spectra there

¹¹In all these cases cancellation of K-theory charges was checked [72] after the results [45] were published, which is the reason why there are a few minor discrepancies between the numbers quoted here and those listed in [45]. All brane configurations in the database of [48] satisfy all K-theory constraints that can be obtained using RCFT probe branes [73].

are huge degeneracies. As already observed in [45], in many cases different boundary state combinations in the same orientifold yield the same spectrum, also for non-chiral states. The origin of these degeneracies is not understood. In the case of MIPFs and orientifolds, permutation degeneracies were removed, and this is usually enough to be left with a set of truly distinct ones. We did not attempt to remove permutation degeneracies for boundary states, but in any case that would not be enough to understand the remaining apparent degeneracies. Furthermore we do not know if the apparent degeneracies are genuine. For example, it is possible that there are slight differences in the set of available instanton branes and their zero-modes, which altogether provides a huge number of parameters that could be compared. But in any case, it is an empirical fact that the number of tadpole free spectra for a given orientifold can exceed the number of *distinct* tadpole free spectra by a factor of one hundred or more. Furthermore, even non-degenerate spectra often have only minor differences in their vector-like states, suggesting that they are nearby points in the open string moduli space. In particular, if one of them satisfies the tadpole conditions for some hidden sector, usually its close relatives have the same property. Even if these degeneracies were fully understood, it is not obvious how to take them into account properly in comparing frequencies of certain features of interest. Therefore statements about this should only be taken as a rough indication.

It turns out that of the 41136 tadpole-free spectra, 31016 come from just one orientifold, and another 9792 from another one. This strongly suggests that two sets of near-degenerate cases dominate the entire sample. We are unable to decide whether this is accidental or whether this should be seen as support for the idea that the presence of discrete symmetries enhances the chance of satisfying the tadpole conditions. Furthermore, we did not attempt to solve the tadpole conditions in those cases where we did not find discrete symmetries. For these reasons, we cannot give a reliable estimate of the likelihood of having both tadpole cancellation and discrete symmetries.

6 Conclusions

In this paper we have studied the appearance of discrete \mathbb{Z}_N gauge symmetries within a large class of RCFT Type II 4d orientifolds with a MSSM-like spectrum. Although interesting for its own sake, our study is motivated by the fact that such discrete symmetries like R-parity or other \mathbb{Z}_N generalizations are necessary to avoid large baryon and/or lepton number violation in the MSSM. Their presence also dictates the possible

signatures of low energy SUSY at the LHC. Thus e.g. if the lightest SUSY particle is not stable and its decay violates lepton number, the search for squarks and gluinos at LHC through R-parity violation channels leads to much weaker mass limits compared to those preserving R-parity.

So a first question is whether indeed such discrete gauge symmetries arise naturally in string compactifications. We have done a systematic search for such symmetries in one of the largest sets available of 4d compactifications with three generations and a MSSM-like structure. These are RCFT Type II orientifold models with modular invariant partition functions based on Gepner models.

One of our most important results is that we have explicitly constructed, for all Gepner orientifolds that can be obtained with simple current MIPFs and all orientifold projections of [58], an integral basis for couplings of axions with all complex branes. This allows us to investigate, under a plausible assumption, whether the $U(1)$'s in a given model are broken completely, or broken to a discrete subgroup. This analysis can easily be performed for any Gepner orientifold, and all discrete subgroups of the full set of $U(1)$'s can be determined in this manner. The integral basis itself may give insight in the geometrical structure underlying to these RCFT models, but we will not explore this issue here.

In this large class of MSSM-like models we did indeed find cases with the appropriate discrete gauge symmetries. The only non-trivial ones were \mathbb{Z}_2 (which turned out to be standard MSSM R-parity) and some other \mathbb{Z}_3 symmetries, not including baryon triality. In models with an additional $U(1)_{B-L}$ there appear symmetries forbidding dimension 5 baryon number violating operators. The finding of these discrete symmetries in string compactifications is, in one hand, good news for theories of low energy SUSY like the MSSM. It shows that symmetries like R-parity, which are imposed in the MSSM in a totally ad-hoc fashion can find a more fundamental origin as discrete remnants $U(1)$ symmetries.

The examples discussed in detail in the previous section display a wide range of positive and negative effects discrete symmetries may have. Basically any effect that was foreseen does indeed occur, with good or bad consequences for dimension four or five B and/or L violating couplings, Yukawa couplings, neutrino masses or the μ -term. In addition one example showed an unexpected feature, namely a discrete \mathbb{Z}_2 symmetry under which fields with both visible and hidden sector quantum numbers are odd. Neutral fields of this kind could provide for new candidates for dark matter.

On the other hand the presence of such symmetries does not seem generic, at least

within the studied class of compactifications, they only appear at a few percent level for MSSM-like configurations. However, it is worthwhile to point out his relative frequency is sizably above what would be obtained from complete randomness, i.e. by considering the axion coupling coefficients as random variables, and computing the probability of them having a given common factor (for instance, for say 10 active axions, the random probability of a \mathbb{Z}_3 can be estimated as $(1/3)^{10} \sim 10^{-5}$). A final remark is that the fraction of RCFT models with discrete symmetries could be enhanced after imposing additional theoretical constraints; for instance we indeed find some indications that the percentage goes up if one considers tadpole-free models. It would certainly be worth studying how often the required discrete symmetries appear in other large classes of string compactifications.

Acknowledgments

We thank F. Marchesano for useful discussions. This work has been partially supported by the grants FPA 2009-09017, FPA 2009-07908 and Consolider-CPAN (CSD2007-00042) from the Spanish Ministry of Economy and Innovation, HEPHACOS-S2009/ESP1473 from the C.A. de Madrid and the contract “UNILHC” PITN-GA-2009-237920 of the European Commission.

A Hodge numbers

In order to help identifying the orientifolds used in the examples, in comparison with geometric data, we specify for each model the Hodge numbers of the corresponding type-II theory as well as their orientifold projection. In this appendix we explain precisely how these number are computed from the RCFT data.

The standard way of computing the spectrum of a Gepner model is by diagonally combining all minimal model and NSR characters. This leads to an $N = 2$ spectrum with certain numbers of hyper multiplets and vector multiplets. These spectra can often be identified with those of ten-dimensional strings compactified on Calabi-Yau manifolds, a statement that can be made precise using the Landau-Ginzburg correspondence [62, 63, 64, 65].

Characters of the chiral algebra (including space-time and world-sheet supersymmetry extensions) of Gepner models that have massless ground states can be expanded as

$$\chi_i(q) = n_{i,s}(s) + n_{i,c}(c) + \text{singlets} + \text{higher order in } q \quad (\text{A.1})$$

In most cases $(n_{i,s}, n_{i,c}) = (1, 0)$ or $(0, 1)$. In the diagonal MIPF these produce spinor-spinor tensor products, leading to $N = 2$ vector multiplets. For example, the tensor product $(3, 3, 3, 3, 3)$ has 4000 characters, of which 202 are relevant for the massless sector: the identity character, 100 characters with $n_s = 1, n_c = 0$, their conjugates, with $n_s = 0, n_c = 1$ and one character with $n_s = n_c = 1$, which is self-conjugate. When diagonally combined with itself, a character with $n_s \neq 0$ and $n_c \neq 0$ yields spinor-anti-spinor products, which lead to $N = 2$ hyper multiplets. For any MIPF, there is an additional hyper multiplet originating from the gravity sector, the square of the identity characters. But characters with both spinors and anti-spinors are rare, and therefore in the diagonal MIPF the number of vector multiplets is usually much larger than the number of hyper multiplets. For example, for the diagonal MIPF of the tensor product $(3, 3, 3, 3, 3)$ we get 101 vector multiplets and 1+1 hyper multiplet.

However, the diagonal MIPF is not the one we use for building orientifolds. The simple current methods we use [58] to compute boundary and crosscap coefficients can only be used for symmetric MIPFs that are obtained by taking a simple current MIPF and multiplying it with the charge conjugation matrix C . In the following we denote a generic MIPF as Z , and a simple current MIPF as Z_S . Multiplication by C does not affect the symmetry of a modular matrix ($(ZC)^T = C^T Z^T = CZ = ZC$ if Z is symmetric, because Z and C commute). There is no general orientifold formalism to deal with Z itself, not even for the special case $Z = 1$. This originates from the

fact that Cardy's work [66], on which the entire formalism is based, dealt with the case $Z = C$, where the Isihibashi states are in one-to-one correspondence with the RCFT primaries. Multiplying the torus partition functions Z with C has the effect of conjugating the space-time spinors in one chiral sector. In the closed string sector this has the effect of interchanging the rôle of vector and hyper-multiplets. The MIPF defined by C itself therefore yields usually a large number of $N = 2$ hyper multiplets and few vector multiplets. For $(3, 3, 3, 3, 3)$, using the MIPF C , one gets 101+1 hyper multiplets and 1 vector multiplet.

Compactification of a ten-dimensional type-II string on a Calabi-Yau manifold with Hodge numbers (h_{11}, h_{21}) yields $h_{21} + 1$ hyper- and h_{11} vector multiplets for type IIA, and h_{21} vector multiplets and $h_{11} + 1$ hyper multiplets for type IIB. Hence if we compare the spectra of the $(3, 3, 3, 3, 3)$ with compactifications on the quintic Calabi-Yau, with $(h_{11}, h_{21}) = (1, 101)$, then the diagonal MIPF matches a compactification of type-IIB on the quintic, while the charge conjugation MIPF matches compactification of type-IIA on the quintic. Since the charge conjugation MIPF is the one to which we apply the orientifold projection, it is natural to use type-IIA language henceforth. The Hodge numbers specified below are therefore obtained in the follow way. From the number of $N = 2$ vector multiplets we get the Hodge number h_{11} , while the number of $N = 2$ hyper multiplets, subtracting one universal one from the gravity sector, gives us the Hodge number h_{21} . We denote the corresponding Calabi-Yau manifold as X .

Note that in ten dimensions Z and ZC are identical (because the spinors are real), and since only symmetric Z can be subject to an RCFT orientifold projection, this implies that only type-IIB strings can be orientifold-projected. In four dimensions, Z and ZC are distinct, and since both are symmetric, both can be orientifold-projected. However, a general formula for boundary and cross caps is only available for $Z_S C$.

Geometrically, the partition function Z can be interpreted either as type-IIB compactified on a manifold X or type-IIA on the mirror of X , while ZC corresponds either to type-IIB on the mirror of X or type-IIA on X . However, in the case of Gepner models the limitation to ZC does not imply that we miss half of the possibilities. It turns out that the set of Gepner simple current MIPFs is closed under mirror symmetry, so that for any hodge pair $(h_{11}, h_{12}) = (p, q)$ there is a MIPF that yields $(h_{11}, h_{12}) = (q, p)$.

In the orientifold theory the diagonal terms in the partition function are subject to a Klein bottle projection. The full partition function has the form

$$\frac{1}{2} \left(\sum_{ij} Z_{ij} \chi_i(\tau) \chi_j(\bar{\tau}) + \sum_i K_i \chi_i(2\tau) \right) \quad (\text{A.2})$$

where $Z = Z_S C$.

The Klein bottle projection works on the states with $i = j$. We can write these as

$$\begin{aligned} & \frac{1}{2} [Z_{ii} \chi_i(\tau) \chi_i(\tau) + K_i \chi_i(2\tau)] = \\ & [Z_{ii} + K_i] \times \left[\frac{1}{2} (\chi_i(\tau) \chi_i(\tau) + \chi_i(2\tau)) \right] \\ & + [Z_{ii} - K_i] \times \left[\frac{1}{2} (\chi_i(\tau) \chi_i(\tau) - \chi_i(2\tau)) \right] \end{aligned}$$

Now we can expand this in ground state spinors

$$\begin{aligned} & \frac{1}{2} (\chi_i \chi_i \pm \chi_i) = \\ & (n_{i,s} n_{i,s} \pm n_{i,s}) (s \otimes s)_S + (n_{i,s} n_{i,s} \mp n_{i,s}) (s \otimes s)_A \\ & + (n_{i,c} n_{i,c} \pm n_{i,c}) (c \otimes c)_S + (n_{i,c} n_{i,c} \mp n_{i,c}) (c \otimes c)_A \\ & + n_{i,c} n_{i,s} s \otimes c \end{aligned}$$

Here “S” means Symmetric and “A” anti-symmetric, and this refers to the sign of the projection obtained from the Klein bottle signs. For fermions, there is an additional spin statistics sign flipping the projection. Therefore the symmetric projection ($K_i = +1$) of an $N = 2$ vector superfield yields an $N = 1$ chiral multiplet, and the anti-symmetric projection ($K_i = -1$) yields an $N = 2$ vector field, even though group-theoretically the vector is contained in the *symmetric* part of the lightcone spinor-spinor tensor product. We denote the corresponding components of h_{11} as h_{11}^+ and h_{11}^- , where the subscript denote the sign of K_i . Hence by h_{11}^- we mean the anti-symmetric Klein bottle projection, yielding the symmetric light-cone spinor-spinor product, which produces a vector boson. Then h_{11}^- is identified with the number of vector bosons in the closed string spectrum, and h_{11}^+ with the number of chiral multiplets. Here we are following the conventions chosen in [45]. In the literature these superscripts are sometimes flipped, so it is better to directly compare the number of vector bosons. Note that h_{21} will get contributions from $(Z_{ii} + K_i)$ and from $(Z_{ii} - K_i)$, so one could split h_{21} into plus and minus contributions, if one wishes to do so. However, both projections of h_{21} yield chiral multiplets, and so we do not make the distinction. In addition to the projected diagonal contributions there are off-diagonal terms which may contribute to both h_{21} and h_{11} . The latter can be split into equal numbers of symmetric and anti-symmetric contributions ($N = 1$ chiral and vector multiplets). The vector multiplets arising from diagonal and off-diagonal terms are added, and this then defines the quantity h_{11}^- . This is subtracted from h_{11} to obtain h_{11}^+ .

Applying this to the example of the quintic, we get the following. The formalism of [58] allows only one Klein bottle projection in this case, which is always symmetric. Therefore the single $N = 2$ vector boson of the type-IIA theory yields a single chiral multiplet and no vector multiplets. The 101+1 $N = 2$ hyper multiplets yield 101+1 chiral multiplets. Each $N = 2$ hyper multiplet contains two $N = 1$ chiral multiplets, but there is also an overall factor $\frac{1}{2}$ in the projection.

The states that can propagate in the transverse channel of the annulus, Moebius strip or Klein bottle (“Ishibashi states”) originate from the terms $Z_{ii^c} \neq 0$ in the partition function. If i represents a massless character, the ground state contains a massless spinor, and then Z_{ii^c} contributes to h_{12} . Massless Ishibashi states give rise to axions and also to tadpole conditions. Hence the smaller h_{21} is, the smaller the number of tadpole conditions, and the smaller the number of axions (note however that there are in general also contributions to h_{12} from terms $Z_{ij}, j \neq i^c$, as well as non-trivial multiplicities $n_{i,s} > 1, n_{i,c} > 1$, which do not give rise to additional axions or tadpole equations). The number of axions determines first of all the likelihood of solving the constraint that Y remains massless. Indeed, in [48] it was observed that standard model configurations are most frequently found for compactifications with small h_{21} , whereas in [47, 45] it was observed that most tadpole solution are found for small h_{21} .¹²

References

- [1] G. R. Farrar and P. Fayet, “Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry,” *Phys. Lett. B* **76** (1978) 575.
- [2] L. E. Ibáñez, G. G. Ross, “Discrete gauge symmetries and the origin of baryon and lepton number conservation in supersymmetric versions of the standard model,” *Nucl. Phys.* **B368** (1992) 3-37.
- [3] A. Martin and G. C. Stavenga, “Non-perturbative proton stability,” arXiv:1110.2188 [hep-ph].

¹²Note that different conventions were used in these papers for the Hodge numbers. In [47] the same convention is used as in the present paper, namely quoting the Hodge numbers of the compactification manifold corresponding to the MIPF Z , whereas in [48] and [47] the Hodge numbers derived from the torus ZC were quoted. This amounts to an interchange of h_{11} and h_{12} .

- [4] Y. Aoki *et al.* [RBC-UKQCD Collaboration], “Proton lifetime bounds from chirally symmetric lattice QCD,” *Phys. Rev. D* **78** (2008) 054505 [arXiv:0806.1031 [hep-lat]].
- [5] T. Banks, N. Seiberg, “Symmetries and Strings in Field Theory and Gravity,” *Phys. Rev.* **D83** (2011) 084019. [arXiv:1011.5120 [hep-th]].
- [6] T. Banks, L. J. Dixon, “Constraints on String Vacua with Space-Time Supersymmetry,” *Nucl. Phys.* **B307** (1988) 93-108.
- [7] L. F. Abbott, M. B. Wise, “Wormholes And Global Symmetries,” *Nucl. Phys.* **B325** (1989) 687.
- [8] S. R. Coleman, K. -M. Lee, “Wormholes Made Without Massless Matter Fields,” *Nucl. Phys.* **B329** (1990) 387.
- [9] R. Kallosh, A. D. Linde, D. A. Linde, L. Susskind, “Gravity and global symmetries,” *Phys. Rev.* **D52** (1995) 912-935. [hep-th/9502069].
- [10] M. G. Alford, F. Wilczek, “Aharonov-Bohm Interaction of Cosmic Strings with Matter,” *Phys. Rev. Lett.* **62** (1989) 1071;
- [11] L. M. Krauss, F. Wilczek, “Discrete Gauge Symmetry in Continuum Theories,” *Phys. Rev. Lett.* **62** (1989) 1221;
- [12] J. Preskill, L. M. Krauss, “Local Discrete Symmetry And Quantum Mechanical Hair,” *Nucl. Phys.* **B341** (1990) 50-100.
- [13] L. E. Ibáñez, G. G. Ross, “Discrete gauge symmetry anomalies,” *Phys. Lett.* **B260** (1991) 291-295.
- [14] T. Banks and M. Dine, “Note on discrete gauge anomalies,” *Phys. Rev. D* **45** (1992) 1424 [arXiv:hep-th/9109045].
- [15] L. E. Ibáñez, “More about discrete gauge anomalies,” *Nucl. Phys. B* **398** (1993) 301 [arXiv:hep-ph/9210211].
- [16] H. K. Dreiner, C. Luhn, M. Thormeier, “What is the discrete gauge symmetry of the MSSM?,” *Phys. Rev.* **D73** (2006) 075007. [hep-ph/0512163].
- [17] K. Kurosawa, N. Maru, T. Yanagida, “Nonanomalous R symmetry in supersymmetric unified theories of quarks and leptons,” *Phys. Lett.* **B512**, 203-210 (2001). [arXiv:hep-ph/0105136 [hep-ph]].

- [18] R. N. Mohapatra, M. Ratz, “Gauged Discrete Symmetries and Proton Stability,” *Phys. Rev.* **D76** (2007) 095003. [arXiv:0707.4070 [hep-ph]].
- [19] T. Araki, T. Kobayashi, J. Kubo, S. Ramos-Sanchez, M. Ratz and P. K. S. Vaudrevange, “(Non-)Abelian discrete anomalies,” *Nucl. Phys. B* **805** (2008) 124 [arXiv:0805.0207 [hep-th]].
- [20] H. M. Lee, S. Raby, M. Ratz, G. G. Ross, R. Schieren, K. Schmidt-Hoberg, P. K. S. Vaudrevange, “A unique Z_4^R symmetry for the MSSM,” *Phys. Lett.* **B694**, 491-495 (2011). [arXiv:1009.0905 [hep-ph]].
- [21] R. Kappl, B. Petersen, S. Raby, M. Ratz, R. Schieren, P. K. S. Vaudrevange, “String-derived MSSM vacua with residual R symmetries,” *Nucl. Phys.* **B847** (2011) 325-349. [arXiv:1012.4574 [hep-th]].
- [22] M. Berasaluce-González, L. E. Ibáñez, P. Soler and A. M. Uranga, “Discrete gauge symmetries in D-brane models,” *JHEP* **1112** (2011) 113 [arXiv:1106.4169 [hep-th]].
- [23] L. E. Ibáñez and A. M. Uranga, “String theory and particle physics: An introduction to string phenomenology,” Cambridge, UK: Univ. Pr. (2012) 673 p.
- [24] L. E. Ibáñez, F. Marchesano, R. Rabadan, “Getting just the standard model at intersecting branes,” *JHEP* **0111** (2001) 002. [hep-th/0105155].
- [25] R. Blumenhagen, M. Cvetič, T. Weigand, “Spacetime instanton corrections in 4D string vacua: The Seesaw mechanism for D-Brane models,” *Nucl. Phys.* **B771** (2007) 113-142. [hep-th/0609191].
- [26] L. E. Ibáñez, A. M. Uranga, “Neutrino Majorana Masses from String Theory Instanton Effects,” *JHEP* **0703** (2007) 052. [hep-th/0609213].
- [27] B. Florea, S. Kachru, J. McGreevy, N. Saulina, “Stringy Instantons and Quiver Gauge Theories,” *JHEP* **0705** (2007) 024. [hep-th/0610003].
- [28] R. Blumenhagen, M. Cvetič, S. Kachru, T. Weigand, “D-Brane Instantons in Type II Orientifolds,” *Ann. Rev. Nucl. Part. Sci.* **59** (2009) 269-296. [arXiv:0902.3251 [hep-th]].
- [29] M. Cvetič, J. Halverson, “TASI Lectures: Particle Physics from Perturbative and Non-perturbative Effects in D-braneworlds,” [arXiv:1101.2907 [hep-th]].

- [30] R. Blumenhagen, M. Cvetič, D. Lust, R. Richter, 2, T. Weigand, “Non-perturbative Yukawa Couplings from String Instantons,” *Phys. Rev. Lett.* **100** (2008) 061602. [arXiv:0707.1871 [hep-th]].
- [31] M. Cvetič, T. Weigand, “Hierarchies from D-brane instantons in globally defined Calabi-Yau orientifolds,” *Phys. Rev. Lett.* **100** (2008) 251601. [arXiv:0711.0209 [hep-th]].
- [32] M. Cvetič, R. Richter, 2, T. Weigand, “Computation of D-brane instanton induced superpotential couplings: Majorana masses from string theory,” *Phys. Rev.* **D76** (2007) 086002. [hep-th/0703028].
- [33] L. E. Ibáñez, R. Richter, 2, “Stringy Instantons and Yukawa Couplings in MSSM-like Orientifold Models,” *JHEP* **0903** (2009) 090. [arXiv:0811.1583 [hep-th]].
- [34] L. E. Ibáñez, A. N. Schellekens, A. M. Uranga, “Instanton Induced Neutrino Majorana Masses in CFT Orientifolds with MSSM-like spectra,” *JHEP* **0706** (2007) 011. [arXiv:0704.1079 [hep-th]]
- [35] C. Beasley and E. Witten, “New instanton effects in supersymmetric QCD,” *JHEP* **0501** (2005) 056 [hep-th/0409149].
- [36] C. Beasley and E. Witten, “New instanton effects in string theory,” *JHEP* **0602** (2006) 060 [hep-th/0512039].
- [37] I. Garcia-Etxebarria, F. Marchesano and A. M. Uranga, “Non-perturbative F-terms across lines of BPS stability,” *JHEP* **0807** (2008) 028 [arXiv:0805.0713 [hep-th]].
- [38] O. DeWolfe, A. Giryavets, S. Kachru and W. Taylor, “Enumerating flux vacua with enhanced symmetries,” *JHEP* **0502** (2005) 037 [hep-th/0411061].
- [39] O. DeWolfe, “Enhanced symmetries in multiparameter flux vacua,” *JHEP* **0510** (2005) 066 [hep-th/0506245].
- [40] M. Dine and Z. Sun, “R symmetries in the landscape,” *JHEP* **0601** (2006) 129 [hep-th/0506246].
- [41] D. Gepner, “Exactly Solvable String Compactifications on Manifolds of $SU(N)$ Holonomy,” *Phys. Lett. B* **199** (1987) 380.

- [42] C. Angelantonj, M. Bianchi, G. Pradisi, A. Sagnotti and Y. S. Stanev, “Comments on Gepner models and type I vacua in string theory,” *Phys. Lett. B* **387** (1996) 743 [ArXiv:hep-th/9607229].
- [43] G. Aldazabal, E. C. Andres, M. Leston and C. Nunez, “*Type IIB orientifolds on Gepner points*,” *JHEP* **0309**, 067 (2003) [ArXiv:hep-th/0307183].
- [44] R. Blumenhagen and T. Weigand, “*Chiral supersymmetric Gepner model orientifolds*,” *JHEP* **0402** (2004) 041 [ArXiv:hep-th/0401148].
- [45] T. P. T. Dijkstra, L. R. Huiszoon, A. N. Schellekens, “Supersymmetric standard model spectra from RCFT orientifolds,” *Nucl. Phys.* **B710** (2005) 3-57. [hep-th/0411129].
- [46] G. Aldazabal, E. C. Andres and J. E. Juknevich, “Particle models from orientifolds at Gepner-orbifold points,” *JHEP* **0405**, 054 (2004) [ArXiv:hep-th/0403262].
- [47] T. P. T. Dijkstra, L. R. Huiszoon, A. N. Schellekens, “Chiral supersymmetric standard model spectra from orientifolds of Gepner models,” *Phys. Lett.* **B609** (2005) 408-417. [hep-th/0403196].
- [48] P. Anastasopoulos, T. P. T. Dijkstra, E. Kiritsis, A. N. Schellekens, “Orientifolds, hypercharge embeddings and the Standard Model,” *Nucl. Phys.* **B759** (2006) 83-146. [hep-th/0605226].
- [49] G. Aldazabal, E. Andres and J. E. Juknevich, “On SUSY standard-like models from orbifolds of $D = 6$ Gepner orientifolds,” *JHEP* **0607** (2006) 039 [arXiv:hep-th/0603217].
- [50] E. Kiritsis, M. Lennek, B. Schellekens, “SU(5) orientifolds, Yukawa couplings, Stringy Instantons and Proton Decay,” *Nucl. Phys.* **B829**, 298-324 (2010). [arXiv:0909.0271 [hep-th]].
- [51] P. Anastasopoulos, G. K. Leontaris, R. Richter, A. N. Schellekens, “SU(5) D-brane realizations, Yukawa couplings and proton stability,” *JHEP* **1012**, 011 (2010). [arXiv:1010.5188 [hep-th]].
- [52] A. N. Schellekens and S. Yankielowicz, “Extended Chiral Algebras and Modular Invariant Partition Functions,” *Nucl. Phys. B* **327** (1989) 673.

- [53] K. A. Intriligator, “Bonus Symmetry In Conformal Field Theory,” Nucl. Phys. B **332** (1990) 541.
- [54] A. N. Schellekens and S. Yankielowicz, “New Modular Invariants For N=2 Tensor Products And Four-dimensional Strings,” Nucl. Phys. B **330** (1990) 103.
- [55] M. Kreuzer and A. N. Schellekens, “Simple currents versus orbifolds with discrete torsion: A Complete classification,” Nucl. Phys. B **411** (1994) 97 [hep-th/9306145].
- [56] G. Pradisi, A. Sagnotti and Y. S. Stanev, “Completeness conditions for boundary operators in 2-D conformal field theory,” Phys. Lett. B **381** (1996) 97 [hep-th/9603097].
- [57] I. Brunner, K. Hori, K. Hosomichi and J. Walcher, “Orientifolds of Gepner models,” JHEP **0702** (2007) 001 [hep-th/0401137].
- [58] J. Fuchs, L. R. Huiszoon, A. N. Schellekens, C. Schweigert and J. Walcher, “Boundaries, crosscaps and simple currents,” Phys. Lett. B **495** (2000) 427 [hep-th/0007174].
- [59] A. N. Schellekens and S. Yankielowicz, “Field Identification Fixed Points In The Coset Construction,” Nucl. Phys. B **334** (1990) 67.
- [60] J. Fuchs, B. Schellekens and C. Schweigert, “From Dynkin diagram symmetries to fixed point structures,” Commun. Math. Phys. **180** (1996) 39 [hep-th/9506135].
- [61] E. Kiritsis, B. Schellekens and M. Tsulaia, “Discriminating MSSM families in (free-field) Gepner Orientifolds,” JHEP **0810** (2008) 106 [arXiv:0809.0083 [hep-th]].
- [62] B. R. Greene, C. Vafa and N. P. Warner, “Calabi-Yau Manifolds and Renormalization Group Flows,” Nucl. Phys. B **324** (1989) 371.
- [63] C. Vafa and N. P. Warner, “Catastrophes and the Classification of Conformal Theories,” Phys. Lett. B **218** (1989) 51.
- [64] W. Lerche, C. Vafa and N. P. Warner, “Chiral Rings in N=2 Superconformal Theories,” Nucl. Phys. B **324** (1989) 427.
- [65] C. Vafa, “String Vacua and Orbifoldized L-G Models,” Mod. Phys. Lett. A **4** (1989) 1169.

- [66] J. L. Cardy, “Boundary Conditions, Fusion Rules and the Verlinde Formula,” Nucl. Phys. B **324** (1989) 581.
- [67] D. -E. Diaconescu and M. R. Douglas, “D-branes on stringy Calabi-Yau manifolds,” hep-th/0006224.
- [68] M. Cvetič, I. Papadimitriou and G. Shiu, “Supersymmetric three family SU(5) grand unified models from type IIA orientifolds with intersecting D6-branes,” Nucl. Phys. B **659** (2003) 193 [Erratum-ibid. B **696** (2004) 298] [ArXiv:hep-th/0212177].
- [69] C. M. Chen, G. V. Kraniotis, V. E. Mayes, D. V. Nanopoulos and J. W. Walker, “A supersymmetric flipped SU(5) intersecting brane world,” Phys. Lett. B **611** (2005) 156 [ArXiv:hep-th/0501182];
F. Gmeiner and M. Stein, “Statistics of SU(5) D-brane models on a type II orientifold,” Phys. Rev. D **73** (2006) 126008 [ArXiv:hep-th/0603019];
M. Cvetič and P. Langacker, “New grand unified models with intersecting D6-branes, neutrino masses, and flipped SU(5),” Nucl. Phys. B **776** (2007) 118 [ArXiv:hep-th/0607238];
I. Antoniadis, A. Kumar and B. Panda, “Supersymmetric SU(5) GUT with Stabilized Moduli,” Nucl. Phys. B **795** (2008) 69 [ArXiv:0709.2799][hep-th]. R. Blumenhagen, V. Braun, T. W. Grimm and T. Weigand, “GUTs in Type IIB Orientifold Compactifications,” Nucl. Phys. B **815** (2009) 1 [ArXiv:0811.2936][hep-th].
- [70] R. Blumenhagen, B. Kors, D. Lust and T. Ott, “The standard model from stable intersecting brane world orbifolds,” Nucl. Phys. B **616** (2001) 3 [ArXiv:hep-th/0107138];
J. R. Ellis, P. Kanti and D. V. Nanopoulos, “Intersecting branes flip SU(5),” Nucl. Phys. B **647** (2002) 235 [ArXiv:hep-th/0206087];
M. Axenides, E. Floratos and C. Kokorelis, “SU(5) unified theories from intersecting branes,” JHEP **0310** (2003) 006 [ArXiv:hep-th/0307255].
- [71] D. Berenstein, “Branes vs. GUTS: Challenges for string inspired phenomenology,” [ArXiv:hep-th/0603103].
- [72] B. Gato-Rivera and A. N. Schellekens, “Remarks on global anomalies in RCFT orientifolds,” Phys. Lett. B **632** (2006) 728 [hep-th/0510074].

- [73] A. M. Uranga, “D-brane probes, RR tadpole cancellation and K theory charge,” Nucl. Phys. **B598** (2001) 225-246. [hep-th/0011048].

Nr	U/S	$U(1)$	ab	ab*	a*c	a*c*	a*d	a*d*	bd*	b*d*	c*d	cd	bc	bc*	Total	\mathbb{Z}_2	\mathbb{Z}_3	Tadp.
			Q	Q	U^c	D^c	D^c	U^c	L	L	E^c	N^c	H_d/L	H_u				
7506	S	1	3	—	3	3	0	0	3	—	3	3	0	0	40590	2152	16	320 (\mathbb{Z}_2)
2751	S	2	3	—	3	3	0	0	3	—	3	3	0	0	869428	0	59808	41136
14704	S	1	3	—	1	2	1	2	0	—	3	0	3	0	380	0	0	0
14062	S	1	3	—	2	2	1	1	2	—	3	1	1	0	304	0	0	0
8745	S	1	3	—	3	2	1	0	4	—	3	2	0	1	92	0	0	0
11196	S	1	3	—	3	4	-1	0	2	—	3	4	1	0	40	0	0	0
10551	U	1	1	2	3	3	0	0	3	0	3	3	0	0	116	0	0	0
1352	U	2	1	2	3	3	0	0	3	0	3	3	0	0	20176	0	1472	0
13058	U	1	1	2	3	3	0	0	1	2	3	3	2	2	68	0	0	0
7573	U	2	1	2	3	3	0	0	1	2	3	3	2	2	14744	0	0	0
16074	U	1	0	3	3	3	0	0	3	0	3	3	3	3	128	0	0	0
7967	U	2	0	3	3	3	0	0	3	0	3	3	3	3	5856	0	0	0
12106	U	1	1	2	3	3	0	0	2	1	3	3	1	1	32	0	0	0
7976	U	2	1	2	3	3	0	0	2	1	3	3	1	1	5764	0	192	0
13844	U	2	1	2	3	3	0	0	0	3	3	3	3	3	1096	0	0	0
14793	U	2	2	1	3	3	0	0	4	-1	3	3	-1	-1	400	0	0	0
13762	U	2	0	3	3	3	0	0	6	-3	3	3	0	0	320	0	0	0
14850	U	2	0	3	3	3	0	0	4	-1	3	3	2	2	96	0	0	0
14792	U	2	0	3	3	3	0	0	0	3	3	3	6	6	32	0	32	0
7488	U	1	1	2	1	2	1	2	0	0	3	0	3	0	2864	0	144	0
13015	U	2	1	2	1	2	1	2	0	0	3	0	3	0	352	0	0	0
18086	U	1	2	1	2	4	-1	1	0	0	3	3	3	0	68	0	0	0
13644	U	1	0	3	1	3	0	2	1	-2	3	1	5	1	8	0	0	0
653	U	1	0	3	0	3	0	3	0	-3	3	0	6	0	4	0	0	0

Table 1: Chiral Spectra of the 24 classes of models. Column 1 specifies the number assigned to the spectrum class in the database, column 2 indicates if the **b**-brane is unitary or symplectic, column 3 lists the number of massless $U(1)$'s, including Y , and the subsequent columns list the chiral brane intersections. The last three columns indicate the total number of spectra in each class that is present in the database, and the total number with a certain discrete symmetry. The last column indicates in how many cases there was also a solution to the tadpole conditions.

Nr	Type	$U(1)$	$A_{\mathbf{a}}$	$\mathbf{a}^*\mathbf{b}$	$\mathbf{a}^*\mathbf{b}^*$	\mathbf{ac}	\mathbf{bc}	\mathbf{bc}^*	A_2	S_2	A_3	S_3	Total	\mathbb{Z}_2	\mathbb{Z}_3	Tadp.
7	UO	1	3	3	—	—	—	—	—	—	—	—	16845	0	0	0
218	UU	2	3	3	0	—	—	—	0	-3	—	—	1049	0	0	0
345	UU	1	3	3	0	—	—	—	0	-3	—	—	1136	18	0	0
742	UU	1	3	2	1	—	—	—	0	-1	—	—	146	0	0	0
18371	UU	1	3	6	-3	—	—	—	0	-9	—	—	12	0	0	0
57	UO	1	3	3	3	3	0	0	0	0	—	—	13402	552	0	0
998	UO	2	3	3	3	3	0	0	0	0	—	—	18890	0	0	0
1000	UUU	3	3	3	3	3	-3	3	0	0	3	0	7276	0	0	0
4004	UUU	2	3	3	3	3	-3	3	0	0	3	0	1706	4	0	0
4316	UUU	2	3	3	3	3	-3	3	0	0	3	0	5236	180	120	0
4324	UUU	1	3	3	3	3	-3	3	0	0	3	0	1278	8	0	0
4325	UUU	1	3	3	3	3	-4	4	0	0	4	1	96	48	0	0

Table 2: Chiral Spectra of the $SU(5)$ GUT models considered here. Here \mathbf{a} denotes the $U(5)$ brane, and \mathbf{b} and \mathbf{c} the additional branes. Column 2 specifies the brane types, and column 3 the number of massless $U(1)$'s including Y (which is embedded entirely in $SU(5)$). The difference between 4004 and 4316 is the embedding of the additional $U(1)$, which is $Y_{\mathbf{b}}$ in nr. 4004 and $Y_{\mathbf{a}} - 5Y_{\mathbf{c}}$ in nr. 4316.